Investigation of Data Dissemination Techniques for Opportunistic Networks

A thesis submitted in partial fulfilment

of the requirement for the degree of Doctor of Philosophy

Halikul Bin Lenando

3 MAY 2012

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

School of Computer Science & Informatics

Declaration

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed .... (candidate)

Date 3 MAY 2012 ....

Statement 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed .... (candidate)

Date 3 MAY 2012 ....

Statement 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed .... (candidate)

Date 3 MAY 2012 ....

Statement 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .... (candidate)

Date 3 MAY 2012 ....

ABSTRACT

An opportunistic network is an infrastructure-less peer to peer network, created between devices that are mobile and wireless enabled. The links between devices are dynamic and often short-lived. Therefore, disseminating data from a source to recipients with a quality of service guarantee and efficiency is a very challenging problem. Furthermore, the interactions between devices are based on opportunity and are dependent on the devices mobility, which have extreme diverse patterns.

The aim of this thesis is to investigate dissemination of data in opportunistic networks. In particular two conflicting objectives are studied: minimising the overhead costs and maximising the information coverage over time. We also take into account the effects of mobility. Extensive computer simulation is developed to explore models for information dissemination and mobility. On top of existing mobility models (i.e. Random Walk, Random, Waypoint and Gauss Markov) a hybrid model is derived from the Random Waypoint and Gauss Markov mobility models. The effect on mobility model on dissemination performance is found to be highly significant. This is based on sensitivity analysis on mobility and node density.

We first consider different baseline push techniques for data dissemination. We propose four different push techniques, namely Pure Push, Greedy, L-Push and Spray and Relay to analyse the impact of different push techniques to the information dissemination performances. The results present different trade-offs between objectives. As a strategy to manage overheads, we consider controlling to which nodes information is pushed to by establishing a social network between devices. A logical social network can be built between mobile devices if they repeatedly see each other, and can be defined in different ways. This is important because it shows how content may potentially flow to devices. We explore the effects of mobility for different definitions of the social network. This shows how different local criteria for defining links in a social network lead to different social structures. Finally we consider the effect of combining the social structure and intelligent push techniques to further improve the data dissemination performance in opportunistic networks. We discover that prioritising pushing over a social network is able to minimise the overhead costs but it introduces a dissemination delay.

ACKNOWLEDGEMENTS

Alhamdulillah, thanks to ALLAH who granted me a dedicated supervisor, Professor R.M Whitaker who guided me to carried out this research. I am deeply indebted to him for his continued support, encouragement and invaluable ideas during this research.

I also like to thank University Malaysia Sarawak (UNIMAS) and Kementerian Pengajian Tinggi (KPT) for supporting my PhD study, especially in terms of finance. Furthermore, I would like to thank Cardiff University, especially the School of Computer Science & Informatics, for providing the research environment. I also want to acknowledge those who have helped and influenced me along the way to complete this research, and especially the researchers in room C2.11 and Cardiff Keluarga batch 2006 - 2011.

I would like to recognize School staff for their valuable contribution to my research project. I also like to thank A. Williams for helping me in my English writing.

Finally I want to thank my family. The encouragement, sacrifice and support from my beloved wife Siti Saniah Kobak and our joyful children, Izzul Husamuddin, Izzatul Nisa, Abdiurrahman and Abid Amaluddeen has been a powerful spiritual source and has given me energy for my PhD. A special thought goes to my parents, my dad Lehando bin Amarsani, my Mum Jumharija Hapuran, and also to my brother and sister for their prayers and never ending support for my success.

CONTENTS

List of Figures

List of Tables

Chapter 1

INTRODUCTION

1.1 Introduction

Inexpensive portable wireless devices now permeate our daily lives. In many case, these devices can directly communicate with each other using communication standards that operate in the unlicensed spectrum band. These make sharing, exchanging, and publishing information directly between wireless devices more easy to perform and this is likely to improve using developments such as opportunistic networks. An opportunistic network is an infrastructure-less peer to peer network, created between devices that are mobile and wireless enabled. This development of technology motivates ways in which information can be efficiently disseminated to recipients. Methods of forwarding information need to minimize consumption of resources as these are limited.

This thesis is devoted to understanding the behavior of information spreading through interaction between portable wireless devices and investigating how to minimize overhead costs, both to avoid unnecessary pushing of content and unnecessary querying of content. An important issue is the mobility of devices. Through different mobility patterns particular pairs of devices may see each other more frequently than others. This means that invisible social structures can naturally exist that we cannot see without monitoring. As in human life these social structures can possibly be used to efficiently convey information. The work in this thesis investigates whether this can be exploited for efficient spreading of information between wireless enabled devices in opportunistic networks.

Because opportunistic networks are infrastructure less, very dynamic in topology, have spontaneous interactions, and have very short live connections, forwarding information from a source to recipients with a quality of service guarantee is very challenging. Furthermore, the interactions between devices are based on opportunity and are dependent on the devices mobility, which may have very diverse patterns. Therefore, to assist this research investigation, we conduct our studies taking into account the effects of different mobility models on information spreading.

Pushing and querying are two major tasks that devices can perform in dissemination. In this work, pushing concerns sending content to a recipient without them requesting it. Querying is requesting information from another device. A device that has not yet discovered information, a query is used for seeking new information. For devices that have information, a query is used for seeking updated information. The updated information is identified based on the age of information. In this thesis, we are focusing on homogenous information only. Both pushing and querying contribute to consumption of resources in mobile communications. However, these tasks are the key to the spreading of information in opportunistic networks and the way they are used affects how quickly information is spread. So, we investigate the processes of push and query and also social structure to understand deeply its effects on information spreading and overhead costs. We aim to understand how the trade-off between performance (information spreading and speed relative to flooding) and overhead costs (resource usage) is affected by push, query and social structure.

1.2 Opportunistic Networking (OPNET)

An OPNET is a disconnected mobile network which has intermittent network connections. The devices communicate directly with each other under the unlicensed spectrum band i.e 802.11g and bluetooth. The communication between nodes occurs on an opportunistic basis. At a single point at time, an end to end path between sender and receiver may not exist. Pairs of nodes may temporarily set up connections between themselves and forward data. Later, the receiver can forward to other nodes in the same way. In the case that the node is disconnected from others, the information that needs to be forwarded is buffered. The information can be forwarded when a potential or the target recipient is within a node's direct contact. Figure 1.1 shows information forwarding in opportunistic networks. At the beginning where t=0, only node a has information. At t=1, both nodes *a* and *b* are in direct contact. Because node *a* has information, information is forwarded to node *b*. Now both nodes *a* and *b* are the information forwarder. At t=2, we can see that all nodes have discovered information as node *a* is in contact with node *c* and node *b* is in contact with node *d*. As we can see from figure 1.1, even though no simultenous connections between node *a* and node *d* existed, but through opportunistic information forwarding, node *a* able to pass information to node *d*.

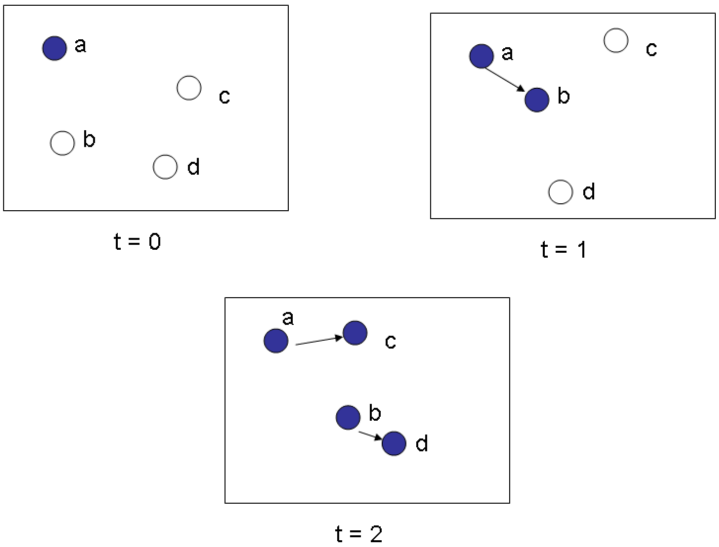


Figure 1.1. Opportunistic Networks

An OPNET is a type of mobile ad hoc network (MANET). Both networks have dynamic reconfiguration of wireless network links. The nodes act both as hosts as well as routers. However in an OPNET a compete path between destination and source is not necessarily available. The nodes utilize the direct contact interaction to choose the best next hop to bring the information close to the recipient. However in MANET, protocols such as AODV [36] and DSR [17] need to establish first a complete route between source and destination before the actual data is forwarded. This is why most of the routing and dissemination protocol of MANETs cannot be used in OPNETs without adjustment.

1.3 Information Provision

Broadcast is the basic way of communication in many systems [38]. A node simply sends data or information to all known neighbours. This approach is used to disseminate information in the situation where no knowledge about network topology is available. Because all nodes are involved in forwarding information, this approach suffers from the broadcast storm problem [33]. Multicast is a type of communication which forwards information simultaneously to a group of nodes (members) only. Multicast is more structured than broadcast. Another type of communication is Unicast which sends information to a single node or host in the network.

In this thesis our concern is Broadcast. In OPNET, the most simplest approach to achieve broadcast is a Flooding approach. For example a node simply sends information to all its known neighbours. Because of this, flooding suffers from high information duplications and uses many resources. Nevertheless, Flooding is still a useful technique to disseminate information because it is simple.

Besides Flooding, Epidemic approaches can be used in OPNET to deliver information. The Epidemic approach was introduced by D.Agrawal [2] to maintain replicated data in a transaction database between servers. This technique is reused by [48] to reduce the duplication problem in the Flooding technique. Instead of forwarding information to all neighbors, the Epidemic selects a neighbour to forward information on a probabilistic basis. The Epidemic approach is a gossip protocol, because gossip spreads information similar to the spread of a virus in a biological community [50].

The Epidemic approach has been used by number of researchers in information provision in opportunistic networks. For example Beaufour [4] and his colleauge use this approach to exchange messages in sensor networks, Juang [19] used epidemic approach in ZebRaNet project to explore the wireless protocols from a power-efficient persepective. Glance [10] used the epidemic approach to disseminate information by harnessing the movements of people.

Context aware routing is also another stream of information provision in OPNET which exploits the user or node information to assist information spreading. HiBOP [5] and CAR [32] are examples of context aware techniques which use device information to route for unicasting. This approach requires a node to maintain a local database to be able to forward information to the next hop. This approach is more complex than the Epidemic approach as it involves many considerations before forwarding information.

Generally, all techniques that are used in wireless communication for information provision have two common basic attributes which are Push and Query. Pushing enables a node to forward information from one node to another. Querying enables a node to send request information to another nodes. To the best of our knowledge, there is no published work that investigates specifically on the effect of Push and Query on information spreading in OPNET. This is the research focus of this thesis. Thus, we establish this research mainly to answer the research question to what extent a Query and a Push technique can minimize the overhead costs and maximize the information spreading delivery in opportunistic networks. In addition, we also investigate to what extent a social structure between devices is useful in disseminating information in OPNET. In general, a relationship between nodes can be made from different interdependencies, such as friendship, knowledge, belief and other elements that make nodes share or exchange things. The social structure has received much attention initiated from the six degrees separation works by Jhon Guare [12] and implemented in small world experiment by Stanley Millgram [31].

1.4 Research purpose and scope

This thesis concentrates on exploring *different ways of pushing and querying with a view to manage information duplications (overhead costs) while trying not to decrease the information dissemination performance (information coverage over time).* Therefore we are seeking *to minimize the overhead costs and maximize the speed of information spreading relatively close to flooding performance in opportunistic networks.* This involves investigating the trade-off in performance verses resource usage.

The research starts by investigating the effects of different mobility models on information spreading. Understanding different mobility models is important because different mobility represent different pattern of nodes mobility. The nodes mobility affect the nodes interaction frequency. Therefore, analyzing the mobility model is necessary in order to further study the information spreading behavior in mobile networks.

The research proceeds by analyzing the effect of push and query on the information spreading. First, we develop a number of different types of simple interaction protocols which have push and query attributes. Then, we test each of the protocols on different mobility models. Furthermore, we also measure the performances on information spreading and overhead costs for each individual protocol.

Chapter 3 and chapter 4 provide a detailed study of the behavior of information spreading in opportunistic networks through the query and push mechanisms. We further modify the push attributes to discover the effect on performance. We apply this techniques using a social structure based on familiarity between nodes. The purpose of introducing the new approaches is to explore whether they can generate high performance trade-off between objectives (resources verses coverage).

1.5 Thesis Organization

After this chapter (introduction), the contributions are outlined as follows:

In **Chapter 2:** related literature to data dissemination in opportunistic networks is discussed. We begin with understanding mobile peer to peer networks and opportunistic networks characteristics as these two concepts are similar. We also present different types of opportunistic network applications that are related to data dissemination. Then we review different techniques for information dissemination in opportunistic networks. At the end of the body of literature, we describe how our research differs from the research that reviewed in the literature.

In **Chapter 3:** The effect of different mobility model on data dissemination is analyzed. We initially compare three different existing mobility models; Random Walk, Random Waypoint and Gauss Markov. However, realizing that Random Waypoint and Gauss Markov possess potential attributes that are useful to represent closely the human mobility, we introduced a new model which called *the D-GM* mobility model. This mobility model is a new hybrid model from both (Random Waypoint & Gauss Markov) mobility models. With the different mobility models, we develop three Key Performance Indicators (KPIs). These are information coverage profile, age profile and update profile. Different mobility models show different effect on the data dissemination behavior.

In **Chapter 4:** The effect of query and push mechanisms on the data dissemination is analyzed. Using the mobility model proposed in chapter 3, chapter 4 focuses on understanding the behavior of information dissemination with different ways of acquiring information through query and push mechanism. This thesis considers Pure Push techniques using existing flooding algorithms. We also introduce three new different basic forwarding techniques based on push and query. They are *Greedy*, *L-Push* and *Spray and Relay*. Each technique has different way of implementing push and query. These are assessed using the KPI in chapter 3.

In **chapter 5:** This chapter investigates formation of a social network based on the nodes interactions. We introduce three ways in which the nodes interact to form a social structure. The techniques that we use to form a social structure are *Average Interaction based*, *Periodic Interaction based* and *Sliding Window based*. In addition we visualize the pattern of social structure for each technique.

In **Chapter 6:** we test the use of a social structure with different flooding techniques. This is carried out using a social structure that formed from frequency of interaction. The social structure is used to choose which nodes to interact with. As we want to investigate the effects of different combinations between social structures and push mechanism on information spreading, we develop and apply a comprehensive set of experiments .

In **chapter 7:** we provide a discussion of our research and its contribution to the study. We also provide how this work can be extended in the future.

Chapter 2

LITERATURE

2.1 Introduction

This chapter discusses the basic concept of opportunistic networks (in section 2.2) and their applications (in section 2.3). The basic information forwarding techniques that allow devices to exchange information are also discussed in section 2.4. This chapter begins by explaining mobile peer to peer networks and opportunistic networks as these two concepts are inter-related to each other. Before we focus on the most important issue in opportunistic networks (i.e message forwarding), we provide a summary on how opportunistic networks may be implemented in the real world. Then we discuss different ways of forwarding information in opportunistic networks. Lastly, we present how our research is differ from the existing forwarding information techniques in section 2.5.

2.2 Mobile Peer to Peer and Opportunistic Networks

The concept of mobile peer to peer networks (MP2P) and opportunistic networks are very similar but explained in different terms. The next subsections provide a general concept of both types of network and summary of works that are related.

2.2.1 Mobile Peer to Peer

Mobile peer to peer networking (MP2P) is a generalisation of Mobile Ad-Hoc Networking (MANET). It can be defined as the interaction between pairs of peer mobile devices for sharing information. MP2P exhibits dynamic network topology, therefore it is hard to maintain end to end connectivity between peers. To facilitate the connection between devices, MP2P may use short range technologies such as Bluetooth, IEEE 802.11g and ZigBee. Routing is less of a priority as compared to a MANET because the primary purpose is pair wise connectivity for sharing. In a MANET, to route data, information is needed about the destination in order to complete the routing task and this can be gained from the end-to-end network structure. Consequently maintaining this end-to-end connectivity is important. In contrast with MP2P, there is no end-to-end connectivity so gaining information to help routing is challenging. However a type of routing for information can be enabled in MP2P by the concept of *store, carry, forward* (SCF) [34]. The SCF technique allows nodes to store content to be forwarded later to a target recipient or population.

In general, we can classify the research in mobile peer to peer into two classes. The first class is using cellular technology for extending the existing wired peer to peer networks to mobile devices. The second class is using short range communication technology such as WiFi, Zigbee and Bluetooth in peer to peer communication.

The cellular peer to peer communication focuses on providing a solution to adapt the existing P2P network infrastructures to the peculiarities of mobile environment which are limited in resources, bandwidth and frequently have disrupted connections. Overlay networks is one of the solutions that are used to extend P2P networks to mobile networks. Dynamo [51] is an example of project that uses proximity information to produce an overlay network structure that similar to the underlying physical network topology. MobiGrid [8] is another project that enables mobile devices to communicate peer to peer with P-Grid [1], where P-Grid is a wired peer-to-peer network. Projects such as Peer-to-Peer Content Sharing Application [30], Mobile Chedar [23] and MOBY [15] use the same concept as Dynamo and MobiGrid project.

In cellular peer to peer communications, a new communication protocol is required to enable a smooth integration between the P2P wired networks and the cellular networks. The main objective of introducing a new protocol is to address how to route information or signaling messages efficiently in a dynamic mobile network topology. Projects like JXTA [20], MPP [41], Generic Engine [13] and JMobipeer [28] are examples of works that propose new protocols to enable communication between mobile devices and wired networks in a peer to peer manner. These protocols enable any mobile devices that are connected to a network to exchange messages independently. MPP, Generic Engine and JMobieer projects not only address the connectivity, they also have proposed an efficient message routing protocol to route information effectively in a dynamic network topology.

Under the second category, the research focuses on the pair wise mobile to mobile communications. The communication is mostly used for collaboration between co-located mobile devices. Data is exchanged based on carry-store-forward mechanism. The carry-store-forward mechanism is where data is carried by a mobile node when a recipient is not in communication range. The challenge to implement this mechanism is to determine an efficient technique for sharing information among mobile devices and to identify which mobile devices can forward or carry information efficiently. Peer2Me [49] provides a set of frameworks for mobile devices to collaborate and to exchange information effectively. It uses the Bluetooth technology as a communication medium for peer-to-peer communication. Because development is focussed on mobile devices, J2ME (Java 2 Platform, Micro Edition) technology is used for application development. This is because J2ME provides a general execution platform for devices that are limited in resources such as mobile phone and PDAs. Peer2Me has similarity with Proem [22] and JMobiPeer [28] in terms of providing a channel for allowing mobile nodes to collaborate in P2P manner.

2.2.2 Opportunistic networks

An opportunistic network has the same attributes as a mobile peer to peer network but it has evolved from a different community. Whereas MP2P has arisen from content sharing and peer-to-peer overlays for file sharing, opportunistic networking has been developed by the networking community as an evolution of delay tolerant networking. Instead of end-to-end connectivity being available at a single point in time, connections may be intermittent and so the path of links between source and recipient is spread over a period of time, maybe with no complete path ever existing between sender and receiver [34]. Therefore, in this situation the nodes have to apply store carry forward techniques in order to deliver the information. Similar to MP2P networks, to minimize the delay, exploiting information from direct interactions is necessary to forward information effectively. Because the concept has evolved from the networking community the assumptions are that routing content to individuals is the primary method to satisfy information requirements.

With the advancement of the communication technology, mobile devices are bundled with multiple services which encourage mobile devices to share resources and services for better utilization. [29] envisages that opportunistic networks will lead to opportunistic computing, where more resources can be shared and exploited. This evolution is mainly driven by the capability of mobile devices to exchange and share heterogeneous (rich resources) services.

Realizing the existence of a wide range of scenarios in mobile opportunistic networks, Chul and Do Young [24] investigate the effect of the heterogeneity of mobile nodes' dynamic contact on forwarding performance in opportunistic networks. The approach used in the paper is based on probabilistic forwarding technique. The challenge of deploying this technique is to determine the optimal probabilistic value for a relay node because this will affects the message delivery time performance. To the best of my knowledge, the latest development on routing in opportunistic network is presented in paper [37] where message delivery is guided using three different forwarding metrics. The forwarding metrics are Group Forwarding Metric (GFM), Probabilistic Forwarding Metric (PFG) and Distance Forwarding Metric (DFM). These forwarding metrics are determined based on the historical node counter information. The challenge to deploy this technique is to select which groups should a mobile device subscribe to in order to receive useful information from other mobile nodes. In [14], an architecture is used to support effective message dissemination in proximity mobile social networks (PMSNs). Through this architecture, users can share information and interact with other users in ad-hoc manner. The architecture also has a database which used by the transmission controller to delete an expired messages to reduce the processing overhead. Not only that, it also decreases the message propagation delay time.

2.3 Opportunistic network applications

Originally, the concept of opportunistic networks was designed to assist military communications and it has now successfully been deployed in a few civil environment where users are more interested to know information that is pertinent to their interests. Opportunistic networks provide an alternative way of communication in areas (for example in a rural area, in a deep forest and in the ocean) where installing network infrastructure is not feasible.

2.3.1 Monitoring animal in sparse environment

Currently there are relatively few applications that have been deployed. The Opportunistic networking idea has been used to facilitate connectivity in order to monitor animal movements using sensor technology through peer to peer communication. SWIM [42] and ZebRaNet [19] are the examples of works that use opportunistic networking concepts to monitor the animal behaviors. The purpose of these works is to study the behaviors of animals mobility through the collection of data and to investigate the behavior of animals.

In SWIM, an infostation model and the peer to peer mobile ad hoc communication concept are combined to study the whales mobility so that the biologists have an idea to preserve them. Whales are equipped with wireless devices to allow information exchange between sensors when they are in contact. The wireless devices on whales are also able to offload information to the infostations which are placed on buoys (floating in the water) when they are in range. Once information is uploaded, the wireless devices memory is cleared. The information from each infostation is sent to terrestrial network or directly to a satellite for further use. Opportunistic forwarding enables the wireless devices to exchange information even though no network infrastructure is available. Determining the best location for placing an infostation is a challenge for data collection and monitoring process.

In the ZebRaNet [19] project, sensors are deployed on zebras which are used as a medium to disseminate information back to the base station. Flooding and history-based protocols are used to exchange information when the wireless sensors on zebras are in range. In the flooding protocol used, a sensor on a zebra floods data to all neighbors when they are discovered. In the history-based protocol used, the devices intelligently select others nodes (zebras) to send information based on prior movement patterns. Each sensor on zebra maintains a level of a hierarchy which is increased when the nodes discover the base station. The hierarchy is used as a metric to guide nodes in selecting which node has a high potential to discover the base station. This work shows that peer to peer networking can be deployed to monitor wildlife without a proper network infrastructure.

2.3.2 As a medium of communication in rural area

Opportunistic networks also have been used in rural areas as a medium of communication, for example in the DarkNet [35] and Saami Nomadic Community (SNC) [9] projects. In the SNC [9] project, opportunistic networks are used to provide basic Internet access to the SNC. Because the SNC always moves from one place to another, mostly in spring and autumn, satellite and wireless communications are used as a medium of communication. However, at the northern regions where the SNC stay, the TCP protocol cannot be used as the nature of data link communication is frequently disconnected. So, bundle protocol is introduced to enable the communication. Bundle protocol [39] is a series of contiguous data blocks that is enough semantic information to equip the application to process data even though there is an individual block may not be ready. Example of applications than can be used to share and distribute information in situation where the communication frequently disrupted are electronic mail, file transfer, and web caching.

In the DarkNet [35] project, the opportunistic networks concept is deployed to enable a small village outside of New Delhi to access Internet services. To enable digital connectivity, data is transmitted over short point-to-point links between a kiosk and portable devices, mounted on buses, motorcycles and bicycles. The approach automatically synchronizes the data from all rural areas when in range with the kiosk using the point-to-point link connections. Through this project, we observe that it is possible to connect people even though continuous connection is not available.

2.3.3 Facilitating connectivity in sparse environments

Opportunistic networks have also been used in facilitating connectivity between nodes in sparse environments. Sparse environments refer to situations where mobile nodes are sparsely distributed and the connections between them are only significant for certain periods of time. Information is disseminated from one node to another through the store-carry-forward system. Because it is not cost effective to install network infrastructures in sparse environments, a three tier architecture is proposed in Data MULEs [40] as a useful model of communication. The top tier is composed of access points which upload the data from MULEs (mobile entities) to a WAN (wide area network). This tier provides reliable connections to central databases which enables the system to synchronize the data that has been collected by MULEs. A MULE is mobile agent that carries and distributes information. The middle tier consists of mobile transport agents which function as bridges to provide connectivity to all tiers. The mobile agents have a large capacity of storage (relative to sensors) and renewable power. The bottom tier consists of wireless sensors which collect information according to their functionality. The sensors communicate with MULEs using short-range radio communication. The key of this architecture is the MULE which provides the overall flexibility and scalability. The same concept is also presented in [53], [54] and [55] in which mobile agents visit the nodes in the sparse networks and distribute information among them.

In forwarding information in sparse networks, the movement of mobile entities are important to facilitate connectivity and to ensure efficiency and reliability of information delivery. A Message Ferry (MF) is a special mobile node that facilitates the connectivity where the nodes are sparsely deployed . Tariq et al [46] propose a predetermined route to improve the reliability of delivery. The predetermined routes consist of two steps. In the first step, the number of stop points and waiting time for each destination are identified. The waiting time is determined based on the probability of MF meeting frequency with nodes. In the second step, all the predetermined points are arranged to find the minimum length to traverse the points. The MF uses the minimum length to traverse the predetermined points.

2.4 Routing in Opportunistic Networks

In opportunistic networks, information exchange protocols between nodes have played an important role for routing in unicast based communication. Unicast based communication [47] is point to point communication or direct communication. Therefore, to establish links between nodes, a dedicated line must be reserved for such connection. Because the nature of opportunistic networks where no communication links can be established between the sender and the receiver (final destination). So, to enable the information to travel over the network using unicast communication based, the nodes have to utilize their interactions with others to enables nodes to identify whether to forward the current information or not. What makes this problem more challenging is that the information exchange protocol should minimize the consumption of resources by maximizing the information delivery time. A number of researchers suggest different solutions to address this issue. From the literature, we categorize the existing routing techniques into three main categories. There are *Flooding Based*, *Epidemic Based* and *Context Based Routing* techniques. The following subsections discuss further each of this category.

2.4.1 Flooding Based Routing

The flooding technique is the simplest forwarding approach possible. It is a broadcast communication approach which forwards a received message or information to all known neighbors whenever possible. When a node encounters other nodes, it simply passes the received information without any considerations. In high density scenarios, a broadcast storm [33] is likely happen when too many nodes are repeatedly sending the same information.

Using flooding to disseminate information is effective when a route from one destination to another is unknown and the network topology is always changing. Under flooding dissemination, at every meeting, nodes exchange information without considering any issues (i.e information duplication). Therefore nodes will discover information very quickly, but when routes are known and predictable, other forwarding techniques are more efficient to reduce the duplication problem and helps in minimizing the unnecessary overhead processing.

In this thesis, we use flooding performance as the benchmark for our study. We compare flooding performance with our forwarding techniques which are introduced in chapter 4 and 5.

2.4.2 Epidemic Based Routing

Epidemic Routing introduced in [48], is an improvement of flooding based techniques. This is a good example of how flooding can be improved to reduce the amount of information duplication. Instead of passing information to all known neighbors [48], it exchanges information index (summary vector) to avoid sending the same information to same nodes. The summary vector contains information that has been seen by the sender. The nodes exchange a summary vector when they are connected in directly. After this exchange, each node is able to determine which information has not been previously seen. This guides a node to request information from another. However, the frequency of information exchange or spread is subject to the size of node's memory. The process of information spread in the epidemic approach is similar to how diseases are spread. First, an infected node that has the information is a carrier to spread the disease. When the carrier encounters other nodes that are not infected (i.e. not received information), it passes the disease to the uninfected node. The nodes stop sending information (spreading disease) when information reaches the destination.

To implement epidemic routing, it is necessary that each message has a unique ID besides the source and destination information. This is to help nodes determine whether the message has been previously seen or not. A number of hops is also useful to control information duplication in networks. For example, when the number of hops allowed is equal to one then the message can only be delivered to its final destination.

Beside the size of memory limitation, nodes have to exchange information and update local information at every meeting. The nodes will be involved in many information exchanges and updates where the frequency of meeting with other nodes is very high. However, the frequent updates can be reduced by avoiding exchange information to the same nodes. This is can be achieved by looking at the history of the nodes interactions.

The epidemic routing technique has been used by number of researchers to disseminate information in the opportunistic networks domain. For example Beaufour [4] and his colleague use this approach in sensor networks, Juang [19] uses the epidemic approach to collect data from sensors though zebra interactions, and Glance [10] uses the epidemic approach to disseminate information by using people as a medium of communication.

2.4.2.1 Probabilistic Routing

Probabilistic Routing, introduced in [27], extends the epidemic routing approach. Each node calculates the delivery predictability of encountered nodes. This probability reflects whether the encountered node is a good forwarder or not. As in epidemic routing, when two nodes are connected, both nodes exchange a summary vector but in probabilistic routing the summary vector contains a delivery predictability value. So, assume that node a and node b are connected and node a wants to forward information to node b. Node a checks whether the delivery predictability value of node b is higher than node a. If this is so, then node a forwards the buffered information to node b.

In probabilistic routing [27], each node established a probabilistic metric called delivery predictability, $P_{(a,b)} \in [0, 1]$, at every node a for each known destination b. The calculation of delivery predictability has three parts; the frequency of node encounter, the age of node encounter, and the transitive property. From the delivery predictability, a node will be able to know the likely probability of message to be delivered to the destination.

2.4.2.2 Spray and Wait Routing

Spray and Wait is a variation of epidemic based routing. It is designed to achieve low delivery delays and energy-efficient in forwarding information in OPNETs. The experimental results in [43] shows that the Spray and Wait approach outperforms the single-copy [45] and multiple-copy [44] routing protocols in terms of total number of transmission and average delivery delay under different traffic loads. A single-copy approach forwards message only to the final recipient, thus, it suffers from high delay delivery time because only a single message is produced for a single destination. For multiple-copy routing, multiple copies of the same message are produced. Moreover, the messages also can be routed independently. This is increases the robustness of the message dissemination to reach the destination.

Spray and Wait routing consists of *spray* and *wait* phases. At the spray phase, the message is injected from a source to a number distinct nodes (relays). The message spread between nodes is initially based on epidemic routing. The number of copies L of message permitted. This controls the overhead by controlling the number of copies of the message.

In Spray and Wait, determining the number of messages injected to the network L is an open issue because it depends on the network size. For the binary spray and wait implementation, a node gives a half of its quota of injected messages to its encountered node, thus the number of messages that can be injected by the sender has is $L = L/2$. When the number of injecting message is equal to one (L=1), a node is only allowed to deliver the message directly to the final destination.

With respect to our work, we have used this popular method as a basis for comparison, integrating the Spray and Wait protocol with query and push mechanisms. In fact, the Spray and Wait approach inspired us to investigate whether using *push* and *query* combinations can be used to reduce the overhead costs of flooding while trying to maintain the information delivery effectiveness as close to flooding technique performance.

2.4.3 Context-Based Routing

Context Based routing exploits user information for routing purposes. Information such as user address, location, and preferences are the examples of information that can be exploited. CAR [32], HiBOP [5] are techniques that are classified under this category. This technique requires a memory space to store information and states of other encountered nodes. This is important for a node to update and to recalculate the potential of its neighbors when forwarding information to other nodes.

In HiBOp [5], a collection of information that describes which community a node is belongs to and its history of social interaction is used to assist information forwarding. This information is exchanged with other nodes during contacts which helps a node to understand its context environment. The current context is useful for evaluating the suitability of node to become a good forwarder. This is critical in the HiBOp approach, therefore each node maintains a History Table (HT) which records the current context and Identify Table (IT) which contain the personal information on a user. This technique is used in chapter 5 to capture the social structure based on the nodes interactions.

2.5 Relevance to our investigation

Our focus in this thesis is the investigation of efficiency and performance for broadcast scenarios. As such, from Section 2.4, the most relevant work to our approach is the epidemic approach. Our work has the same objective as the epidemic approach which is to maximize the delivery performance and minimize the use of resources [48]. However the works does not explicitly address the effects of push and query, nor does it use the social structure in data dissemination. Push is passing information from one node to another node directly. Query is a message send from node to another to discover new information. In the original epidemic approach, when nodes are in range, a vector (containing the index of all information kept in the node's memory) is pushed. On receiving the vector, the nodes compare it with the local vector information. Then the nodes acknowledge the vector message by sending the missing index information that is found during the comparison process. Upon receiving the acknowledgment, the sender sends the missing information as a reply to the acknowledgment. This is a very specific protocol that uses push and query of high level information in a very specific manner.

In our investigation, we address query and push in a generalised manner, dealing with homogeneous information. A single type of information allows us to focus only on query and push overheads. Our work is also inspired by the flooding technique, which acts as a benchmark to consider performance. It is a very simple technique that allows maximum spreading of data but at high cost. It has massive utilization of resources which is not efficient for opportunistic networks as device resources are limited. Based on the literature, it is only the Spray and Wait [43], [45], [44] approaches that use purely push techniques in disseminating information. These techniques limit the number of messages to be pushed as to control information duplication. In comparison to this thesis, a social network structure is used to avoid duplication to particular nodes through a logical structure that defined based on nodes mobility patterns. Within the work, we develop an understanding of the different push techniques, how they can be combined with querying others for data and how they impact upon performance. This contribution is valuable as it has to the best of our knowledge not been previously carried out.

In our investigation, no acknowledgement is used because we are only dealing with homogeneous information, so indexing different types of information is no longer needed because all nodes are only interested in a single type of information. We use different techniques of push and query mechanism to spread information and investigate the input of these. We investigate the range of push size (which in the Spray and Wait approach is limited based on the number of nodes) to observe its effect on data dissemination performance. Moreover, we also introduce a quota concept which clarifies in the Spray and Wait [43] approach regarding controlling the number of messages that can be injected on the network by a node. We named this approach as Spray and Relay as presented in chapter 4.

Exploring the impact of push and query combinations without any intelligent assistance is vital research because it helps us to understand the impact of push and query techniques behavior on data dissemination. Further more, it also paves the way to improve the push technique and motivates the design of an intelligent dissemination based approach. In [27], [11], [16], and [18], the past contact history has been used to help nodes to disseminate information effectively. Recently, nodes history of interactions has also been used in [5] to form a social structure. The goal of the social structure is to help nodes to forward information effectively to the right nodes at maximum performance. In relation to our work, we define clearly the concept of the social structure constructed from the nodes history interactions (i.e. the frequency of nodes seeing each other in a given period of time). Moreover, we also investigate different ways of constructing social structures in chapter 5.

From chapter 4, we observe that push protocols have great impact on information spreading performance in opportunistic networks. This motivates us to further investigate the possibility of achieving the best information dissemination performance by just using the push protocol. As a result, we present different push techniques where the goal is to maximize the delivery time and minimize the duplication problem. Not only that, we also integrate our push techniques with the social structure using a construction that we develop in chapter 5. As far as we can see, this research is the first investigation that attempts to address the push and social structure combinations to maximize the delivery time performance and minimize the overhead costs in opportunistic networks.

Chapter 3

MOBILITY MODEL AND DATA DISSEMINATION

3.1 Introduction

Mobility is important in our work because it is through mobility that nodes gain the opportunity to disseminate data to other peers. It is often the case that authors use a single mobility model. This is not good because the choice of mobility model affects the results. Therefore, this research deploys different mobility models to study the sensitivity of different mobility models on data dissemination performance in opportunistic networks.

The purpose of this chapter is to understand the data dissemination behaviour over time of various levels of mobile node density and different types of common mobility models. We use four different mobile models. Three mobility models are taken from [7] i.e. Random Walk, Random Waypoint, and Gauss Markov and we develop another mobility model which is a combination of Random Waypoint model and Gauss Markov model. We named the model is a *Directed Gauss Markov (D-GM)* mobility model. The rest of the chapter is organized into five subsequent sections. In section 3.2, we briefly discuss the motivation underpinning this chapter. Because the result here is based on computer simulation, we have described our simulation model in Section 3.3. The experiments and results are presented in Section 3.4. Section 3.5, provides discussion with regard to this chapter's objective.

3.2 Motivation

The characteristics of opportunistic or Mobile Peer to Peer (MP2P) make simulation modelling a very useful tool for understanding the behaviour of these networks. For example, MP2P networks has dynamic topology. So, in order to understand MP2P networks, we need to monitor each of the nodes communication which is very complex and time consuming. However, through simulation, repeatable scenarios with different settings can be evaluated very fast. Moreover, via refinement of settings, it helps to deepen understanding of how changes (parameters) impact upon the MP2P networks.

The movement patterns of mobile node play a vital role in data dissemination in mobile peer to peer (MP2P) networking. For example, a user's mobility behaviour affects the number of peers discovered which in turn increases the opportunity for exchange messages (artifacts). The more frequently nodes are discovered, the higher the chance of spreading messages quickly.

Besides the movement patterns, the density of users also plays an important role in understanding data dissemination. For example, in a very busy city, the opportunity of meeting other users is high. Thus, the chance of exchanging messages between users is also high. This is contributes to make data disseminate quickly.

Using different mobility movements and different levels of node density are thus essential input for our simulation and the investigation of the data dissemination behaviour. We include two random models (Random Walk and Random Waypoint) [7] in order to simulate a commonly used mobility pattern. The Random Walk model simulates a situation where a person who randomly walk around in a mall. A Random Waypoint model simulates human movement similar to Random Walk but stop at particular location before continue to move to other location in a mall.

Another model included in our simulation is the Gauss Markov [7] movement model. This movement is categorized as dependent movement, where the calculation of next movement is based on the previous direction and speed. Interestingly, this movement overcomes the sharp turning issue present in the Random model. Sharp turning is when a node is stopped and turns to the different direction. This model simulates a human movement pattern that has a particular destination and no straight path exists. So, to reach the destination, alternative routes which actually deviate from the straight path is used.

Under the Gauss Markov model, a general direction is determined by mean direction ($\bar{d}$). The initial value of $\bar{d}$ for each node is determined depends on the located of a node in the simulation area. For example if the node is located at location x=5 and y=5, then the initial direction is 45 degree. This is to force nodes move to the center of the simulation area.

In order to understand the effect of the number of nodes in particular place on information dissemination, we model nodes that have location and movement based on the defined mobility models. We also increase the number of nodes by 15 for every simulation by taking into consideration the existing nodes in each simulation. The aim is to simulate how the different levels of density actually influence the data dissemination behaviour.

3.3 Model

For purposes of just understanding mobility, the simulation is developed to simulate the data exchange between mobile nodes which appear in a given range (e.g: 30m). The data exchange occurs only if a new artifact is discovered between peers. The peer discovery process is opportunistic, in that the mobile node has no knowledge who (mobile nodes), what (kind of artifact) and when (in which particular time) it will discover another peer. To simplify the simulation, we used a single data (artifact) type for all nodes which originated from one source (info-station). The artifact has an age which indicates how long the artifact has been in the network. The age of artifact is incremental by one following the simulation step. This is important because it is through the age we can identify the update frequency information in the networks. The artifact that originates from the information source always considered as a fresh information (i.e. age =0). The following subsections explain in detail the simulation components.

3.3.1 Transmission

We assume that the transmission technology applied has basic characteristics for personal area networking (e.g. Bluetooth). The parameters used for this experiment are given

Table 3.1. Global Parameters

Parameters	**Setting**
Channel Bit Rate	10 Kbps
Discover Success rate	0.95
Transmission Success rate	0.95
Channel setup time	0.5
Meta-data size	0.1kb
Artifact size	3kb
Transmission range	30 meters
Simulation time step	0.1 seconds
Simulation duration	15 Minutes
Region size	500m x 500m

in Table 3.1. The parameters are chosen to have a clear and systematic experimental investigation. We assume a message size of 3 kb, allowing a text message of around 1,500 to 3,000 characters (depending on encoding). Since we only have to transmit 3.1 kb data (artifact and meta data) over the channel, 10 kbps is sufficient for the data transfer between nodes. The connection is based on peer to peer communication so the nodes are only allowed to have one connection at time, hence we assume a data rate transmission success of 95%. We also assume that the success rate of discovering other nodes is 95%. To setup a communication channel requires a small period of time; for our experiment we give a conservative estimate of 0.5 seconds for this purpose. Random selection of the discovered partner node is assumed and each node only maintains one link at a time.

3.3.2 Information exchange protocol

We apply a fully opportunistic protocol for artifact exchange. The protocol is greedy in the sense that when a node is not engaged in peer-to-peer interaction, it is engaged in peer discovery. Note that this is not a resource efficient protocol and it is unlikely to be appropriate in practice but appropriate to model a specific protocol. We are modelling it here merely to scope the possible performance in terms of data dissemination quality that could be achieved using different mobility models.

Once a connection has been established and channel set-up is completed, the pair of nodes exchange meta-data which describes the age of their current artifact or the absence of an artifact. At this point, each node can determine whether it is required to transmit or receive an update of the artifact. In the case of a node is engaged with an information station (i.e. information source), the node artifact is updated directly after connection is established without performing the meta-data exchange process. An artifact update between nodes occurs if either

1. One of the nodes has no artifact, in which case a copy of the artifact is transmitted
2. Both nodes have artifacts of different ages, in which case the older artifact is updated.

The communication between peers is bi-directional. However, no acknowledgment is performed and a 95% transmission success rate is assumed for both data and meta-data, as long as the nodes are in range for the required data transmission.

3.3.3 Simulation components

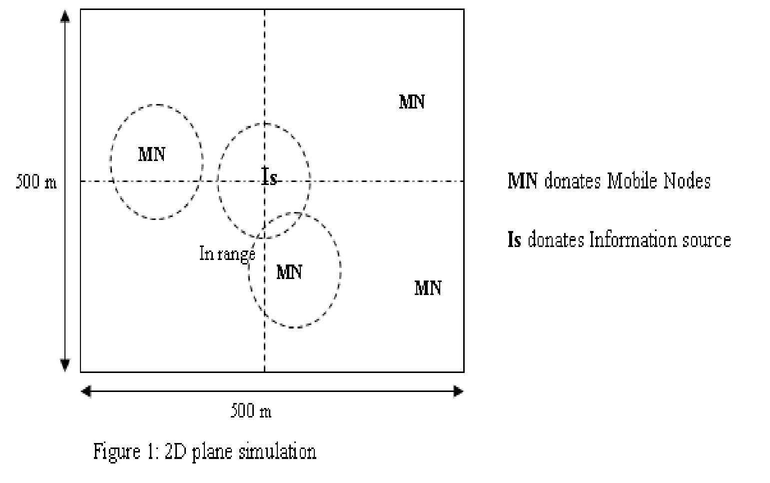


Figure 3.1. 2D plane simulation

Figure 3.1 shows the plane and entities in the simulation. The following outline the entities used in the simulation:

- Mobile node (MN)- MN is a portable device (Pda, Mobile Phone, Laptop) which is embedded with wireless communication such as Bluetooth, zigbee, or Wi-Fi. It is capable of exchanging messages with its peer. Moreover, it also operates as a carrier to spread the artifact.
- Information source (IS) - Information source is a device which is embedded with wireless communication such as Bluetooth, Zigbee, or Wi-Fi and sensor, that is

situated in a particular place. It only produces the latest information and sends information but not receive information from its peers.

- Simulation region - Simulation region represents the size of an area (in square meters) in which the node can be positioned, moved or discovered. It is also used to limit the nodes' movement, i.e. the node that moves beyond the simulation region area will bounce off depending on which mobility models are in used. Each point in the simulation region is referenced using x and y coordinates.

- Time steps - A time step is measured in seconds. In every time step, each node moves to a new location, based on the mobility model.

- Mobility model - Mobility model dictates the node's movements according to the models that already described. The movement to a new location is calculated at a time step, based on the speed and the direction.

- Artifact -Artifact is a piece of information that travels between nodes. It consists of time (when it was produce)and ID.

- Assumptions about artifact storage - Each mobile node has the capability of storing an artifact.

We use a discrete time-step model for the simulation. At each time step the following elements are updated:

- Node location - location (x, y) for each node

- Neighbour information - the latest information of the closest neighbour is updated for every node

- Artifact age - This is increased using time steps. Age of artifact is increased by one for every time step. The latest artifact will overwrite the old artifact.

3.3.4 The mobility model movement

Nodes in the simulation model move according to Random Walk, Random Waypoint and Gauss Markov mobility models as previously mentioned. Random Walk and Random

Waypoint are categorized as movement independent, which means next movement is not influenced by its previous attributes. In contrast with Gauss Markov movement, the next location is calculated based on the previous speed and direction. The nodes are placed randomly at the beginning of the simulation. For all models, when the nodes hit the boundary, they will bounce back towards center of the simulation area. The following subsections describe the four movements used in the simulation including their parameters.

3.3.4.1 Random Walk

Initially, each mobile node is given two random parameters, direction and speed. The node travels along the trajectory for a fix time interval. Before the node moves to a new location, a new random direction and speed is given. The speed is uniformly distributed between ***[4 km/h, 8 km/h]***. In our simulation, we assumed that nodes changed direction and speed every 30 seconds as to mimic the randomness of the human who walks randomly in open space. We called this time value an interval. For each interval, a new direction and speed will be assigned randomly. Note that, because the Random Walk model selects direction randomly, there is a possibility that a node heading to its previous position. Therefore, with a small time interval, a node is expected to be moving in a small area of coverage, even more restricted than other mobility models.

3.3.4.2 Random Waypoint

Random waypoint is an extension of Random walk. This model introduces a pause time at each interval time, where nodes stay at a location for a certain time (pause). Before the node moves to a new location, a new random direction or destination is given at a speed uniformly distributed between ***[minSpeed, maxSpeed]***. In the simulation, the speed range is chosen between [4 km/h, 8 km/h] which suits the pedestrians' walking speed. Furthermore, we used 30 seconds as the waiting time at each destination to force the nodes to stop at the particular location. Note that, the waiting time here is not fixed and be changed.

3.3.4.3 Gauss Markov

Initially, each mobile node is assigned a random speed and direction. At fix interval time n, new value of speed and direction is calculated based on the following formula [7].

$$s_n = \alpha s_{n-1} + (1-\alpha)\bar{s} + \sqrt{(1-\alpha^2)}s_{x_{n-1}} \tag{3.3.1}$$

$$d_n = \alpha d_{n-1} + (1-\alpha)\bar{d} + \sqrt{(1-\alpha^2)}d_{x_{n-1}} \tag{3.3.2}$$

Here s_n , d_n is the new speed and direction at time interval n. α ,where $0 \leq \alpha \leq 1$, is the tuning parameter used to vary the randomness. $\bar{s}$ and $\bar{d}$ are constant values representing the mean value of speed and direction. The s_x and d_x are the value taken from Gaussian Distribution with the mean equal to zero and standard deviation is equal to one. The $\bar{d}$ is changed over time depending to the edge proximity of the node current location.

At each time interval the next location is calculated based on the equation 3.3.3 and 3.3.4.

$$x_n = x_{n-1} + s_{n-1}\cos(d_{n-1}) \tag{3.3.3}$$

$$y_n = y_{n-1} + s_{n-1}\sin(d_{n-1}) \tag{3.3.4}$$

where (x_n,y_n) is the new location at interval n and (x_{n-1},y_{n-1}) is the previous location at interval $n-1$. The (s_{n-1},d_{n-1}) are the previous speed and direction before moving to the interval n.

To ensure the nodes remain in the simulation area, the mean ($\bar{d}$) of nodes is changed based on the nodes location as shown in figure 3.2 which taken from [7]. For example if a node near to the left edge of the simulation area, the $\bar{d}$ is set to 0 degree. This forces a node to move towards the center. A node is considered near to the edge when the distance between the node and the edge is 20 meters.

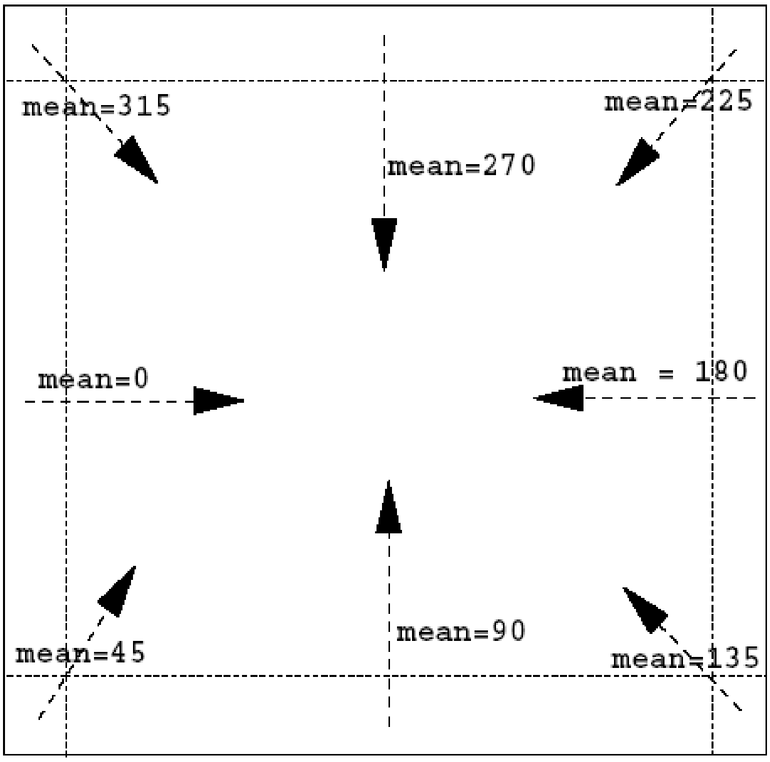


Figure 3.2. Change of Mean Angle Near to Edge (in degree)

3.3.4.4 Directed Gauss Markov (D-GM)

A D-GM is a combination of two mobile models i.e. Random Waypoint model and Gauss Markov. The movement of a node is based on Gauss Markov mobility model. Each node has predefined stops which depends on a node's group. A D-GM takes a Random Waypoint stop attribute where nodes stop at every predefined stop. Each stop has its own predefined stop duration which is defined at the beginning of the simulation.

Under the Gauss Markov model, mean direction ($\bar{d}$) of a node is determine based on which simulation boundary a node is near to. The $\bar{d}$ is determine the general direction of which a node is heading to. However in D-GM a node's general direction is set based on a node's group. Each group has predefined list of destinations to be visited and each group has different priority value for different places.

The mean direction of a node is depends on the list of destination. Once a node has selects it's general direction, the Gauss Markov mobility model is used to calculate the next location. Once a node reaches at the target destination, a node has to stop based on the location's pause time for certain destination. The algorithm of this model movement is presented in Algorithm 1. Figure 3.3 shows the destinations that used in the D-GM model.

Algorithm 1 D-GM mobility model for 9000 simulation steps for a single node

```
initial mean direction (according to a node's group)
initial mean speed
initial current direction
initial current speed
update the node location
pauseTime = 30
for simulationstep = 1 to 9000 do
  if the node has reached its predefined destination then
    if pauseTime > 0 then
      decrease pauseTime by 1
    else
      select new destination
      pauseTime = 30 time steps
    end if
  else
    update the node's new location
    calculate the node's new direction
    calculate the node's new speed
  end if
end for
```

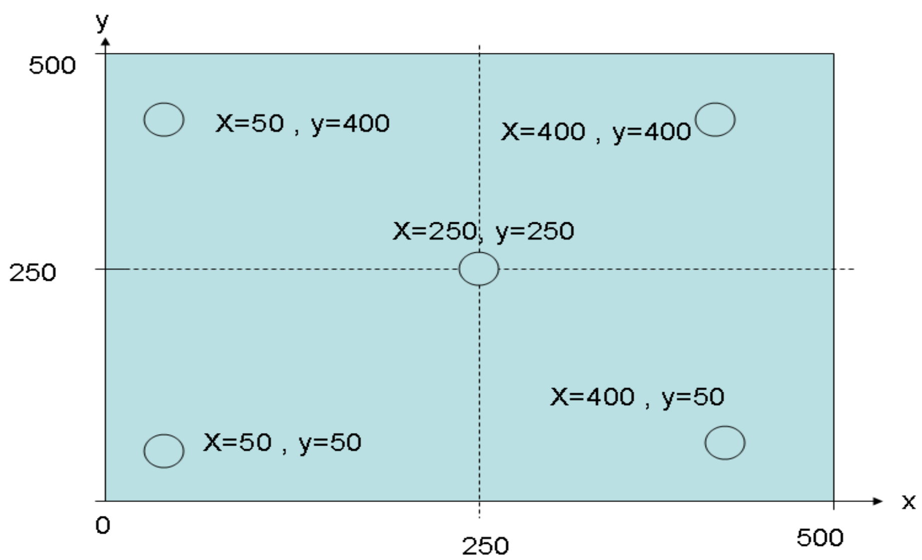


Figure 3.3. Destination coordinates for D-GM model

Table 3.2. Priority sets for each group of nodes

Group	x=50,y=50	x=250,y=250	x=400,y=50	x=50,y=400	x=400,y=400
Group 1	0.2	0.2	0.2	0.2	0.2
Group 2	0.4	0.2	0.1	0.2	0.1
Group 3	0.2	0.3	0.1	0.4	0.1

3.3.5 Test problem scenario

We adopt a 500m x 500m region, which is sufficiently large to represent a large store, small shopping mall or city plaza. We assume that there is a single information source that has a fixed location in the middle of the region, with the same transmission range as mobile nodes (30 meters). We vary the density of mobile nodes in this region using 10, 25, 50, 75 and 100 mobile nodes. Random starting positions are allocated such that, if S_i denotes the starting positions of the test set of i nodes then:

$$S_{10} \subset S_{25} \subset S_{50} \subset S_{75} \subset S_{100} \tag{3.3.5}$$

For all experiments performed, we consider the behaviour of node movement and information acquisition based on 100 random trials. Each trial represents 15 minutes of system operation with the artifact only being held by the source node at the start of each trial.

For the D-GM model, we divide the nodes into three groups. The first group consists of node ID from 1 until 33, the second group from 34 until 66 and the third group from 67 until 100. For each group we have three sets of probabilities to determine the next destination of the nodes. Table 3.2 shows the priority sets that are used for the respective groups.

3.3.6 Performance metrics

In order to assess the quality of information that peers can maintain, we define a range of metrics or Key Performance Indicators (KPI's) that helps us to understand system behaviour. The metrics outline is as follows:

- Artifact distribution

  This metric assesses how quickly the information spreads to all nodes over the simulation period. Rapid dissemination is preferred.

- Profile of artifact age

  This metric assesses the age of an artifact throughout all the simulation steps. Only artifact that produce by the information source has age = 0. The age of artifact is increased by 1 at every time steps. A distribution which is positively skewed towards a smaller artifact age is preferred.

- Number of updates

  This metric measures the number of updates. An updates is replacing an old artifact with an updated artifact which was been discovered through nodes interactions. Frequent updating is preferred but this also requires a high level of resources.

- Total update profile

  This metric measures the average of different level of degrees where degree is defined as a frequency of artifact update between nodes. Involving all nodes to participate at every interaction is preferred.

- Spatial Node Distribution

  This metric is used to examine the distribution of nodes in the simulation region within 15 minutes. The size of simulation region is 500 x 500 square meters. The simulation region is divided into a number of smaller areas which we call *cells*, each of which is 20 meters x 20 meters.

3.4 Experiments and Results

The experiments conducted in this section are organized based on five different KPI's as mentioned in Section 3.3.6. The main purpose of the experiments is to examine the effect of different mobility models and node density on the data dissemination behaviours. The duration of simulation for every experiment is 9000 simulation time steps (15 minutes) and each simulation is repeated 100 times (trials) with different random seed generation (to avoid bias in the result).

3.4.1 Artifact Distribution

In this section, we present the result of artifact distribution behaviour over time using various levels of node density and different types of mobility models.

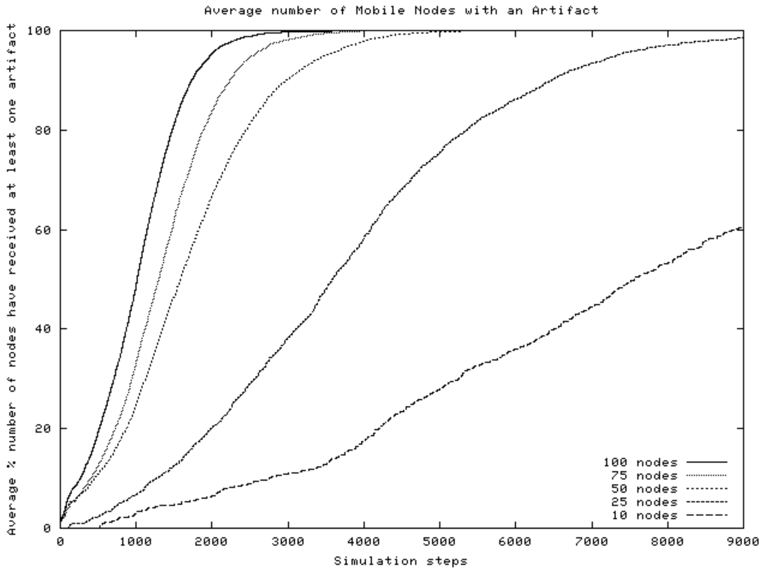


Figure 3.4. Average of mobile nodes received an artifact using Random Walk over 100 trials

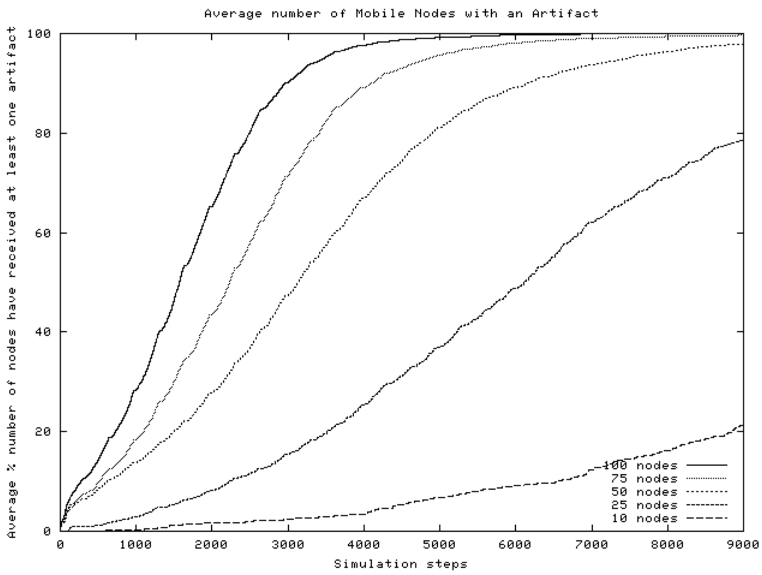


Figure 3.5. Average of mobile nodes received an artifact using Random Waypoint

Figures (3.4, 3.5, 3.6, 3.7) show the average number of mobile nodes received an artifact within 15 minutes of the simulation time using Random Walk, Random Waypoint, Gauss Markov and D-GM mobility models. Each line indicates the levels of density results for

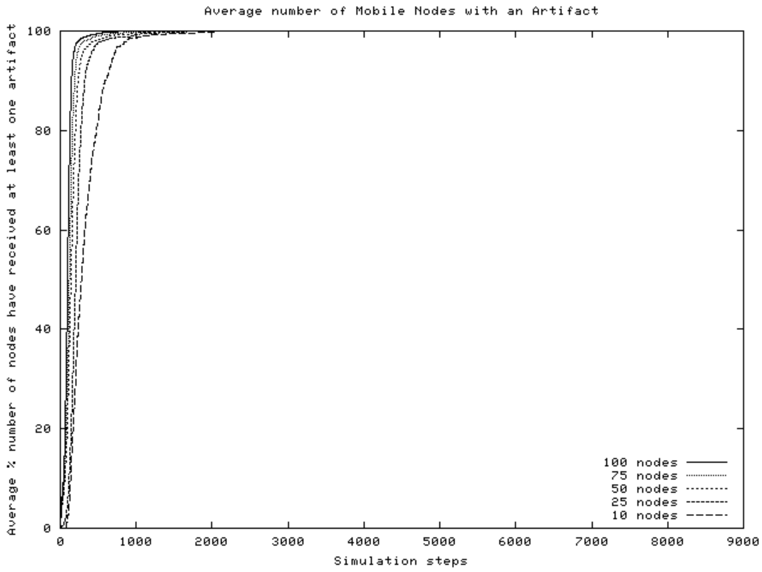


Figure 3.6. Average of mobile nodes received an artifact using Gauss Markov Model

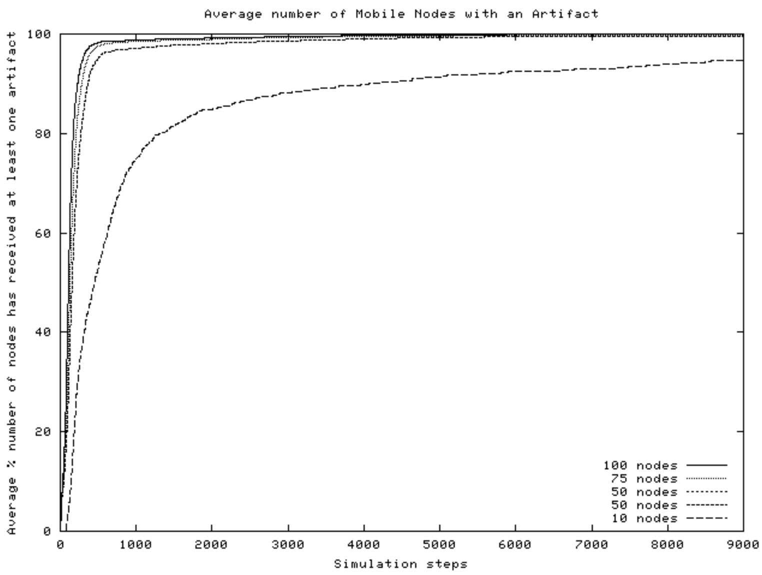


Figure 3.7. Average of mobile nodes received an artifact using D-GM Model

10, 25, 50, 75 and 100 nodes respectively. The statistics of the graphs are given in Table 3.3, 3.4, 3.5 and 3.6.

From Tables 3.3, 3.4, 3.5, and 3.6, it can be observed that the artifact distribution is sensitive to the node density. Increasing the node density also increases the opportunity for nodes to discover an artifact very quickly. This is because more nodes are potentially involved in forwarding an artifact.

Mobility model also affects an artifact distribution performances. For example, Gauss

Markov mobility model distributes an artifact very quickly compared to Random Walk and Random Waypoint mobility models. This is can be seen from Figure 3.6. This is because mobility model determines the frequency of meeting with different nodes. The higher the meeting frequency with different nodes, the higher the chance of nodes discovering an artifact. Looking at the artifact distribution specifically in Figure 3.5, the line with node density = 10 nodes (which is the lowest line in the graph), only has 21.3 % nodes that have an artifact after 15 minutes simulation time. Whereas with the same node density and the same simulation time, Gauss Markov mobility model has achieved 100 percents nodes that have received an artifact. This is because Gauss Markov model creates more chances for nodes to discover an artifact at the center after moving to the edge of the simulation. Moreover, the information source that located at the center, makes easier for nodes to discover an artifact very quickly.

Under Random Walk model, nodes move randomly (i.e., unpredictable) and because of that there is possibility that the nodes move back towards its previous location. This is reduces the chances of node to meet different nodes which indirectly limits the chances of node to discover an artifact. As we expected, Random Walk model takes longer period of time to discover an artifact compared to Gauss Markov model. This is can be observed by comparing the number of nodes received an artifact in percentile for both Tables 3.3 and 3.5.

Random Waypoint model is also random based movement. This model forces the nodes to pause for a 30 seconds at the particular point (destination). This limits the nodes movement and also confine the opportunity of nodes to be updated by other nodes. Thus, there are a high number of nodes without artifact found under this model. This is can seen from the Figure 3.5 where the average of nodes that received an artifact over the time is increases slowly. This is indicates not many nodes discover artifact under this model.

D-GM model gives better performance compared to Random Walk and Random Waypoint models in terms of interactions and artifact discovery. This is because D-GM creates a high chance for nodes to discover each other in one of the destination list. Moreover, because of the movement of nodes are based on the destination list, there is a high potential

Table 3.3. Number of nodes received an artifact using Random Walk

Node density	*number of node received an artifact in percentile*				*Average (mean)*	*Standard Deviation*
	25	50	75	100		
10	8.20	23.50	42.50	60.7	25.57	19.22
25	24.48	68.20	92.20	98.44	58.38	34.66
50	75.20	99.30	100	100	81.35	30.17
75	90.55	99.97	100	100	85.29	28.52
100	97.77	100	100	100	88.65	25.49

Table 3.4. Number of nodes received an artifact using Random Waypoint

Node density	*number of node received an artifact in percentile*				*Average (mean)*	*Standard Deviation*
	25	50	75	100		
10	1.80	5.30	10.80	21.30	6.93	6.09
25	10.12	31.28	58.84	78.48	34.493	25.74
50	32.34	74.44	92.86	97.92	63.07	32.12
75	50.81	93.11	98.89	99.64	74.11	32.12
100	74.19	98.76	99.89	99.97	81.55	29.22

for a node to interact with different nodes. As compared to the GM mobility model, the D-GM mobility model has lower performance. This is because of the nodes that using the D-GM mobility model pause when they reach at destination point. This limits the chance of nodes discovering each other and disseminate information. In contrast, nodes under the GM mobility model keep moving until the end of the simulation. This continues the dissemination of artifacts through the simulation.

Table 3.5. Number of nodes received an artifact using Gauss Markov

Node density	*number of node received an artifact in percentile*				*Average (mean)*	*Standard Deviation*
	25	50	75	100		
10	77.70	98.50	99.40	100	96.28	15.43
25	96.96	98.88	99.68	100	97.51	13.34
50	97.96	99.42	99.90	100	98.27	10.74
75	98.86	99.68	99.92	100	98.57	9.96
100	99.29	99.95	100	100	98.89	8.92

Table 3.6. Number of nodes received an artifact using D-GM

Node density	*number of node received an artifact in percentile*				*Average (mean)*	*Standard Deviation*
	25	50	75	100		
10	85.90	90.60	92.90	94.70	85.18	16.22
25	94.20	96.20	97.52	98.20	92.65	14.02
50	98.26	99.20	99.48	99.60	96.88	11.00
75	99.08	99.56	99.76	99.87	97.77	10.12
100	99.26	99.77	99.90	99.94	98.22	9.32

3.4.2 Artifact Age Profile

The age of an artifact can be used as to represent the interaction frequency of a node. A low value of an artifact age indicates high frequent update. So, this section presents experimental results on the average of an artifact age. We assume that artifact i leaves the infostation at time step x. The age of an artifact i at time step t is therefore $t - x$.

We are concerned with the age of all artifacts at the last time step of a simulation. We create a profile of ages of artifact at last time step that include all the simulation trials that we have conducted (Figures 3.8 - 3.13). The statistics of each experiment can be found in Table 3.7, 3.8 and 3.9. The skewness value is used to estimate the distribution of artifact ages. A skewness characterized the degree of asymmetry of a distribution relative to mean. A positive skewness indicates a distribution of artifacts with an asymmetric tail extending toward more positive values, whereas a negative skewness indicates a distribution with an asymmetric tail extending toward more negative values.

Increasing the node density also increases the chance of nodes meeting with different nodes. This cause more fresh artifacts can be found in a high node density. This is because an artifact has high chances to be updated through a frequent interaction between nodes. This is can be further observed by looking at the statistics in Tables 3.7, 3.8, 3.9 and 3.10 where the lower node density (node density =10) has a high value skewness compared to the high node density (node density =100). This is indicates that there are many old

Table 3.7. Profile of artifact age using Random Walk model

Node density	***Artifact Age***			
	Mean	Skewness	Standard Deviation	Median
10	25.57	82.04	64.57	26.7
25	68.60	50.54	121.39	35.12
50	115.65	20.81	218.38	17.98
75	162.39	11.45	347.52	11.99
100	210.92	7.85	509.90	8.99

Table 3.8. Profile of artifact age using Random Waypoint model

Node density	***Artifact Age***			
	Mean	Skewness	Standard Deviation	Median
10	6.93	92.61	54.51	6.00
25	37.27	82.35	119.99	29.34
50	81.70	57.16	195.39	24.52
75	128.57	36.65	304.78	17.99
100	177.99	23.44	446.86	14.23

Table 3.9. Profile of artifact age using Gauss Markov model

Node density	***Artifact Age***			
	Mean	Skewness	Standard Deviation	Median
10	96.28	32.11	619.67	0.10
25	136.02	20.67	1102.73	0.04
50	166.29	17.00	1767.55	0.02
75	209.43	18.08	2723.49	0.013
100	255.95	19.91	3847.10	0.01

Table 3.10. Profile of artifact age using D-GM model

Node density	***Artifact Age***			
	Mean	Skewness	Standard Deviation	Median
10	8.52	49.60	16.14	4.96
25	23.17	33.84	48.51	10.315
50	71.74	28.02	186.38	27.29
75	145.10	23.94	439.16	49.02
100	243.32	20.92	837.52	75.49

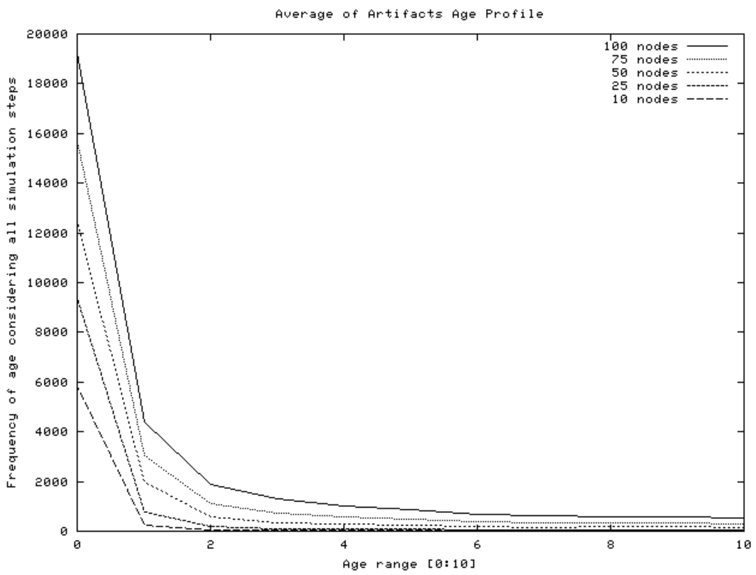


Figure 3.8. Age Profile Frequency using Random Walk Model with age range[0,10]

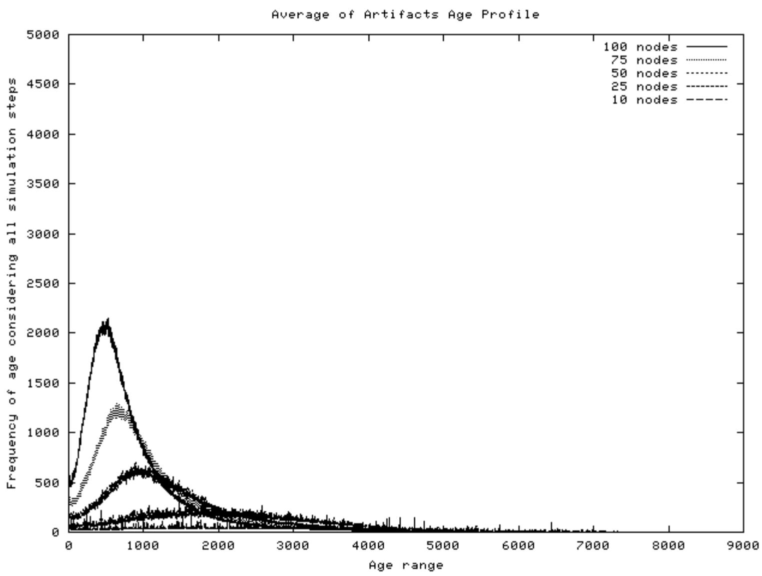


Figure 3.9. Age Profile Frequency using Random Walk Model with age range[0,9000]

artifacts (not updated) in the lower node density. This is not helping for information dissemination.

From the mobility models perspective, Gauss Markov outperforms the other mobility models in keeping an artifact frequently updated. This is can be observed by looking at Table 3.9 where there are large number of low age artifact (age median=0.01) found within 15 minutes simulation. This is because Gauss Markov model directs the nodes to move to the center when they are close to the edge of simulation. This creates opportunity for

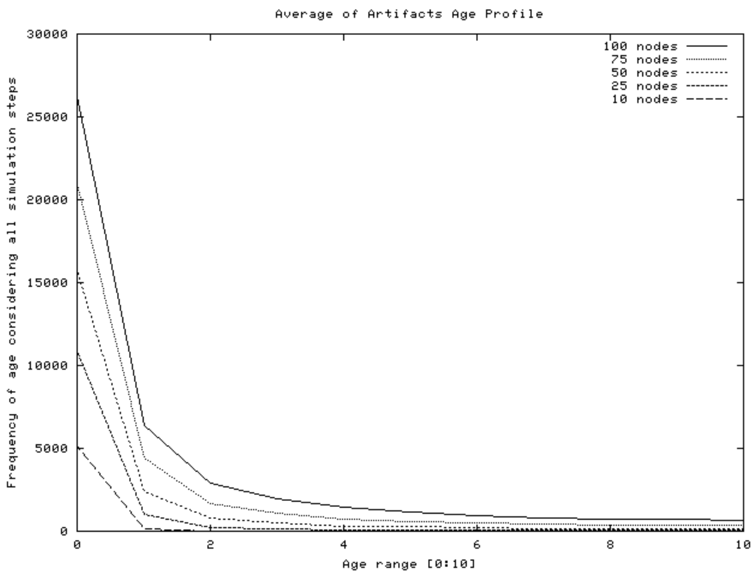


Figure 3.10. Age Profile Frequency using Random Waypoint Model with age range[0,10]

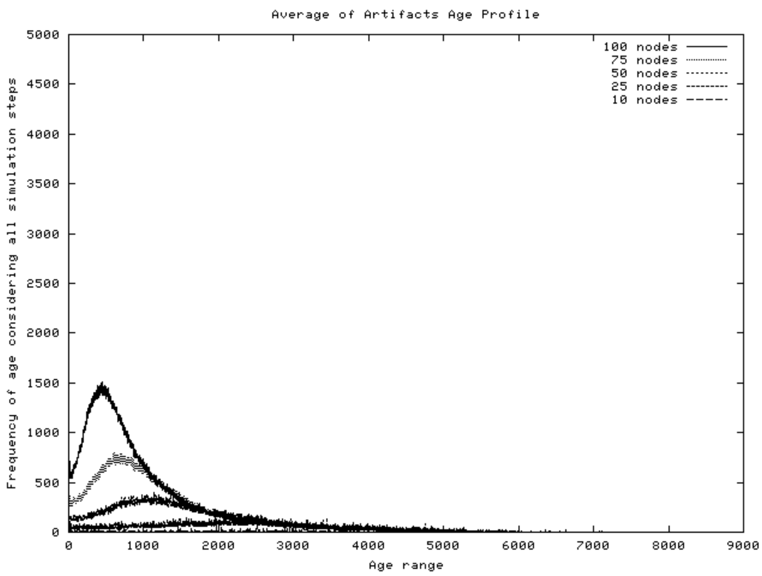


Figure 3.11. Age Profile Frequency using Random Waypoint Model with age range[0,9000]

nodes to interact with different nodes which also helps the nodes to discover an updated artifact. Moreover, the information source that located at the center also keeps the nodes' artifact always updated.

For the D-GM model, the profile of age is close to the Gauss Markov's model profile age. As the density of node increases the number of low age artifact increased. This is due to the fact that more potential forwarders information are available in a high density

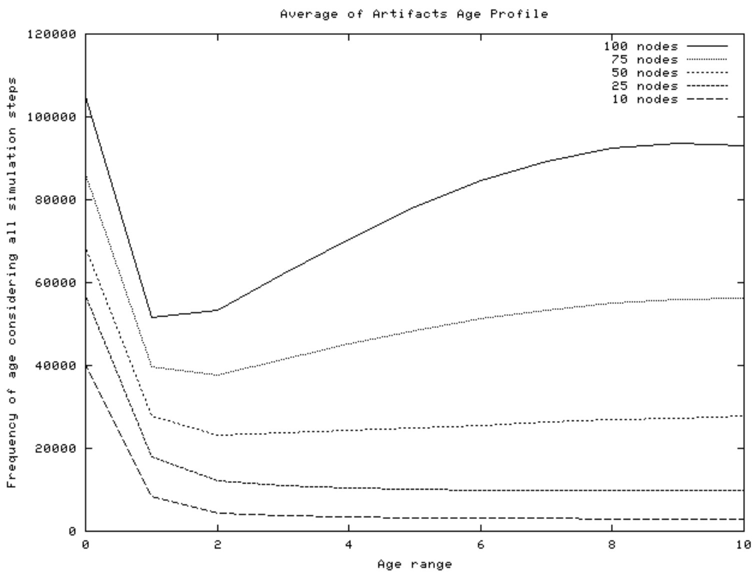


Figure 3.12. Age Profile Frequency using Gauss Markov Model with age range [0,10]

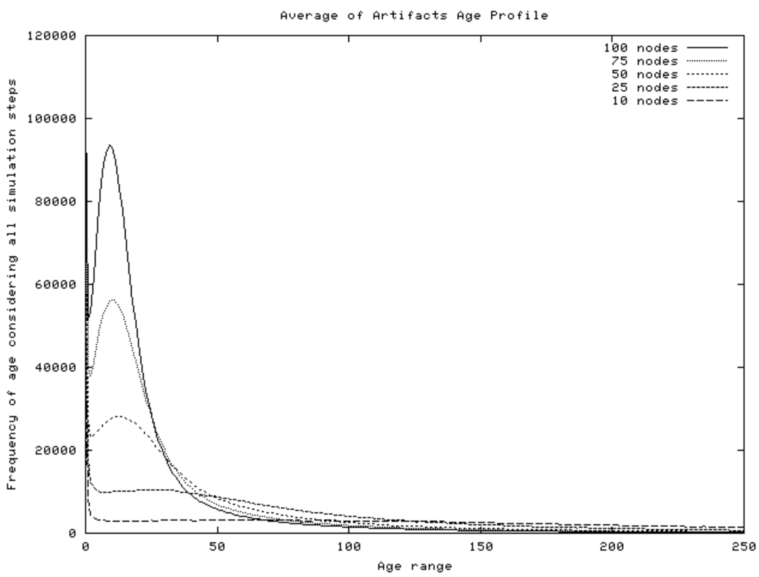


Figure 3.13. Age Profile Frequency using Gauss Markov Model with age range[0,250]

node. Because the nodes are moving towards specified destination in D-GM model, the possibility of meeting other nodes are very high. Moreover, one of the stop of nodes is at the center where the information source is located. Therefore it is reasonable that D-GM has more low age artifact compared to Random Walk and Random Waypoint model as shown in Figure 3.15.

As we expected, Random Walk model has more older artifact age compare to Gauss Markov model. This is because not many nodes have a chance to received a fresh artifact

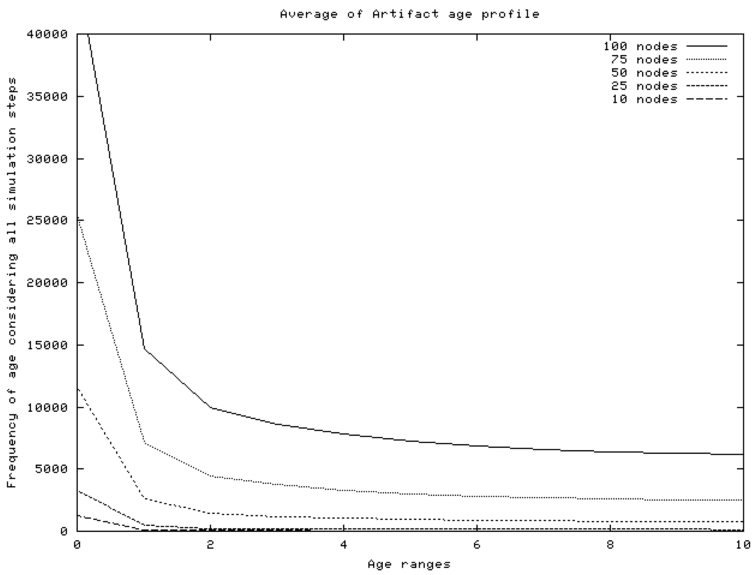


Figure 3.14. Age Profile Frequency using D-GM Model with age range [0,10]

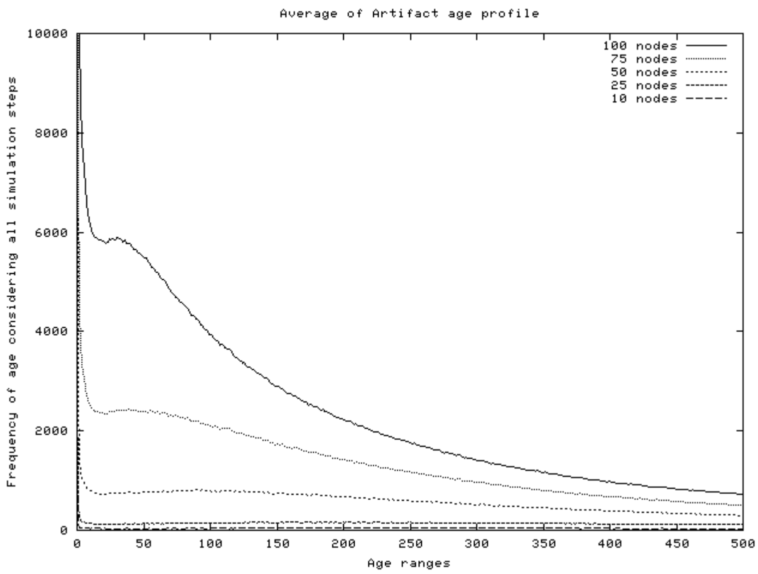


Figure 3.15. Age Profile Frequency using D-GM Model with age range[0,500]

from the infostation. In the case there is a node received a fresh information from the center, as this node travel to the edge of simulation, the artifact is getting older. This situation causes more old artifact found under the Random Walk model. This case is similar to Random Waypoint model. However, with the pause attribute, Random Waypoint has the capability to keep longer the fresh artifact when the nodes are pause in the center area. This is why in the Figure 3.10 Random Waypoint has more fresh artifact compared to Random Walk model (Figure 3.8) when node density is set to 100 even though the

Table 3.11. Number of updates using Random Walk

Node density	***Frequency update in percentile***				***Average of frequency update***
	25	50	75	100	
10	1.1	2.5	3.5	5.0	2.29
25	5.0	10.44	12.04	15.08	8.57
50	17.52	23.66	24.56	26.78	19.96
75	33.29	38.48	39.16	41.59	33.10
100	50.18	54.15	54.76	57.07	47.70

Random Waypoint has limited movement compared to Random Walk model.

3.4.3 Number of updates

A high number of updates is likely contribute to rapid data dissemination. But, more updates will produce high overheads. Hence, in this section we investigate the effect of different node density and different mobility models on the number of artifact updates between nodes. Figure 3.18, 3.16 and 3.17 show the update activity for different mobility models. From the figures, we can see that jitter appear on each graph. This is because, at each time step there is quite large variations in the number of updates that may occur.

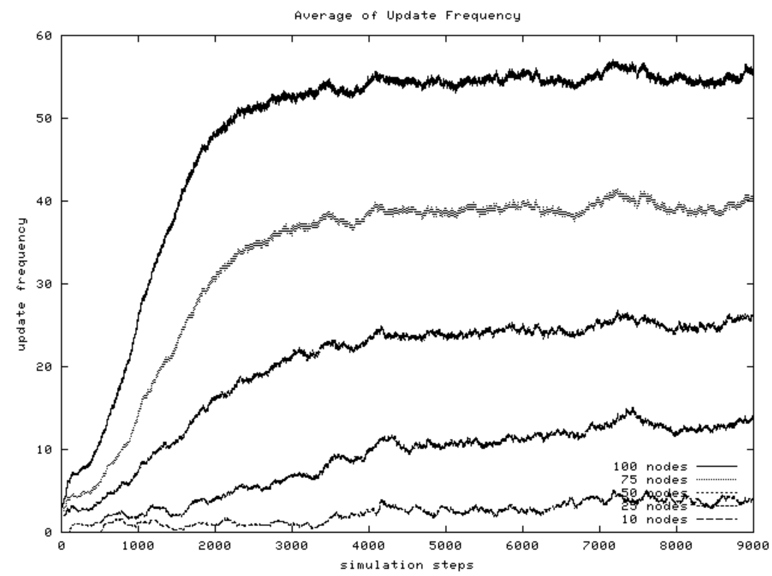


Figure 3.16. Number of updates profile using Random Walk model

From Figures 3.16, 3.17, 3.18 and 3.19 we can observe that the number of updates increased over the simulation time except for the D-GM model. This is because the opportunity of meeting different nodes is different at every time step. So, there is no guarantee that the number of updates at every step is the same. From the figures, we

Table 3.12. Number of updates using Random Waypoint

Node density	*Frequency update in percentile*				*Average of frequency update*
	25	50	75	100	
10	0.4	1.0	1.6	3.5	0.98
25	2.48	5.76	11.08	14.84	6.53
50	8.74	18.18	24.66	28.12	16.70
75	19.33	34.4	39.77	43.23	29.16
100	34.25	51.52	55.5	58.23	43.32

Table 3.13. Number of updates using Gauss Markov

Node density	*Frequency update in percentile*				*Average of frequency update*
	25	50	75	100	
10	13.6	14.3	14.8	17.4	13.87
25	28.28	28.84	29.32	31.16	28.19
50	42.7	43.04	43.52	45.32	42.36
75	58.72	59.08	59.44	60.92	58.16
100	74.66	75.03	75.37	76.73	73.99

Table 3.14. Number of updates using D-GM

Node density	*Frequency update in percentile*				*Average of frequency update*
	25	50	75	100	
10	0.16	0.21	0.29	0.9	0.24
25	0.78	0.98	1.49	4.01	1.24
50	3.27	4.1	5.92	14.77	4.99
75	9.05	10.88	14.91	33.74	12.77
100	18.48	22.39	29.08	61.02	25.26

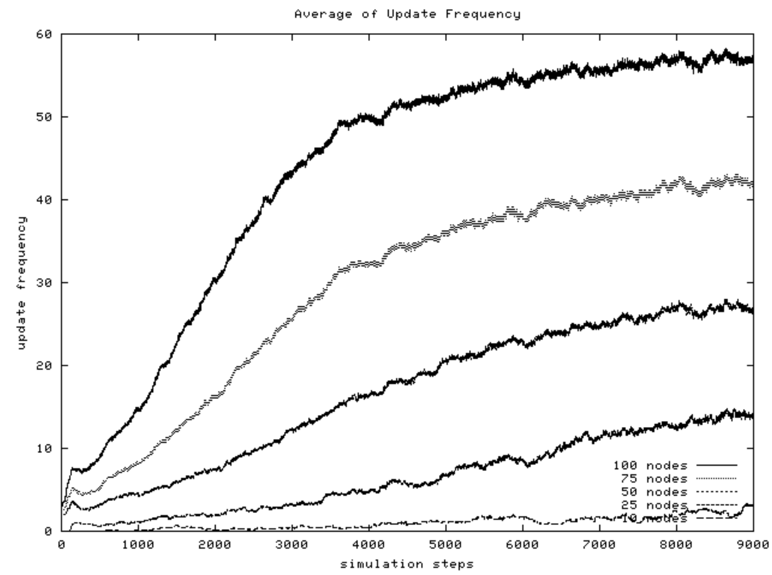


Figure 3.17. Number of updates profile using Random Walk model

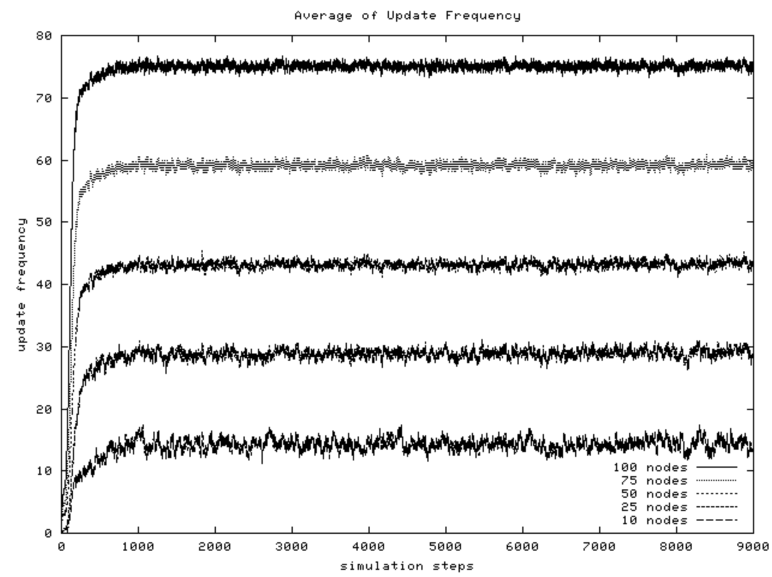


Figure 3.18. Number of updates profile using Gauss Markov model

can see also when the node density is increased, the number of updates is also increased. This is due to the fact that more nodes are involved in updating an artifact (in high node density).

Besides the node density, the mobility models also affect the number of updates. As we can see from Table 3.11, 3.12, 3.13 and 3.14 different mobility models have different average number of updates. Comparing all the mobility models, we found that Gauss Markov model has a high number of updates i.e 73.99 % in 100 node density. This is

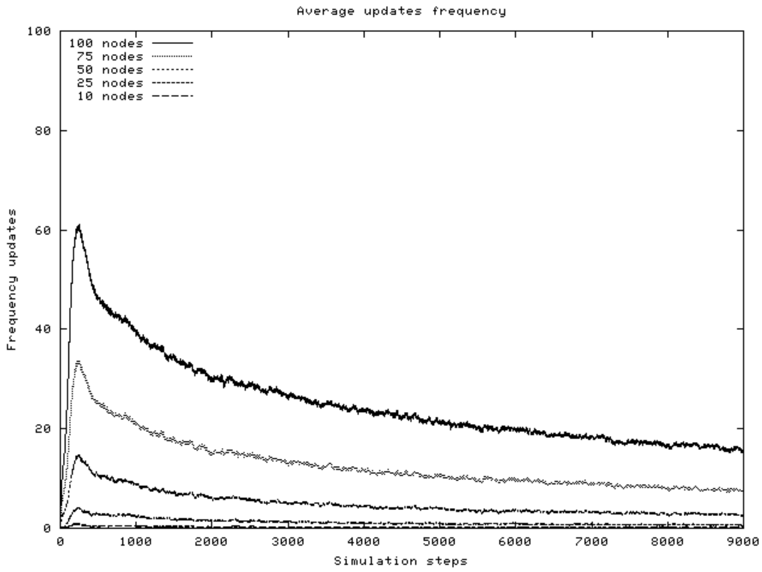


Figure 3.19. Number of updates profile using D-GM model

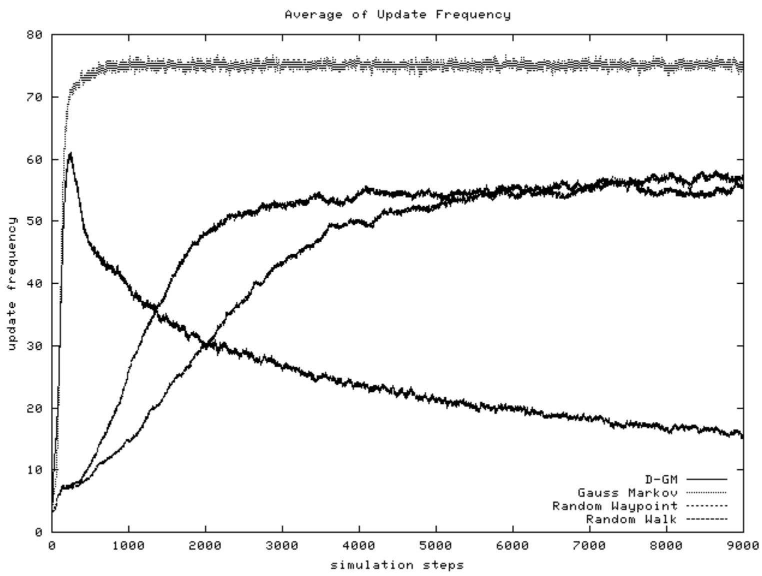


Figure 3.20. A comparison of number of updates profile using different mobility models for 100 nodes density

because Gauss Markov model creates a high chance for nodes to meet each other at the center of the simulation. Moreover, the nodes that travel via the center also have a high chance to receive an artifact from the infostation.

From Figure 3.20, it can be seen that the Gauss Markov model has a stable average number of updates compared to the Random Walk model, Random Waypoint model and D-GM model. This is because the next location of a node in Gauss Markov model is estimated based on the node current location. Therefore, the possibility of a node maintaining its current connection when moving to the new location is higher as compared to Random Walk and Random Waypoint model. The Random Walk and Random Waypoint are using random assignment to determine the next location of a node. So, to maintain the current connection in the location is very hard. The D-GM model has a different pattern of number of updates from other mobility models. The number of updates drops as the simulation time increases. This is because the nodes can be disconnected from others as they have different direction of destination. Moreover different destinations have different pause times which also contributes to the drop in the number of updates.

3.4.4 Total updates profile

This KPI is useful to understand how the nodes participated in the interactions. The ideal situation for this KPI is to have an equal frequency of interactions for all nodes.

The interactions between nodes are recorded throughout the simulation time. From the interactions information, we classified them into different categories of interaction frequency [0, 2000] which identified as *degree* . A degree equal to 0 shows that there are nodes that are never involved in any interactions.

Figures 3.21, 3.22, 3.23 and 3.24 show the degree of the nodes total updates using different mobility models. As we can see from the figures, different mobility models have different degree of nodes total updates. This is because different models have different possibility of meeting with different nodes. Looking at the degree of interactions using Random Waypoint in Figure 3.22, the degree of total updates is more dispersed compared to the degree of total updates using the Gauss Markov (Figure 3.23) and the D-GM (Figure 3.24). This shows that the Gauss Markov and D-GM are better in terms of providing a chance for all nodes to be involved in any interactions as compared to Random Waypoint. This is because the Gauss Markov and the D-GM model have high total updates with a low degree of interactions. Whereas the Random Waypoint has number of high degree total updates which indicates there are many nodes not frequently involved in interactions. The statistics of the figures are tabulated in Table 3.15.

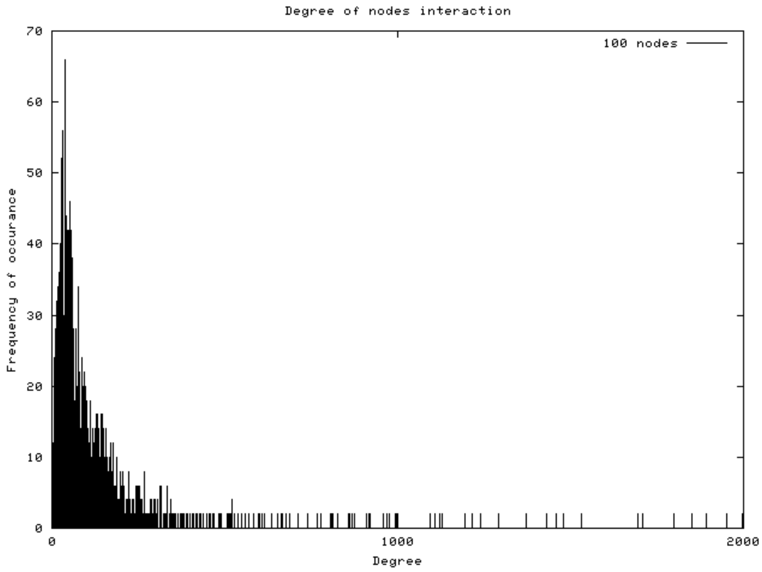


Figure 3.21. Degree of total updates profile using Random Walk Model

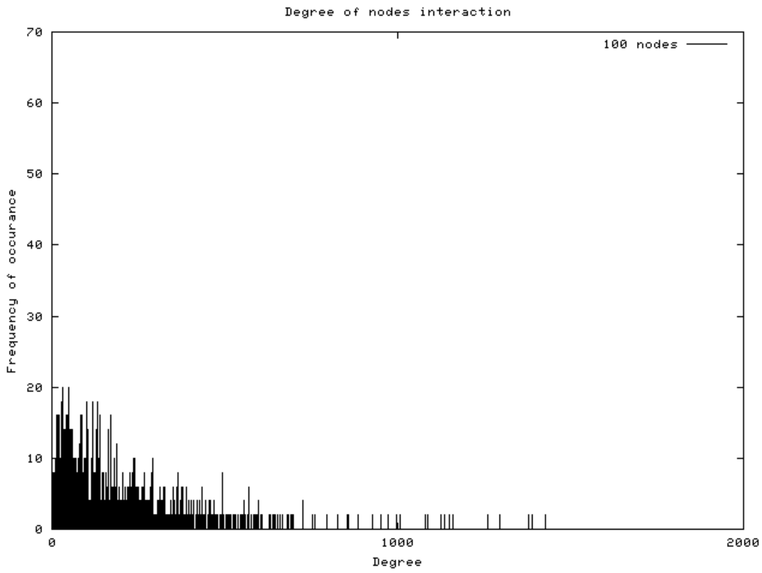


Figure 3.22. Degree of total updates profile using Random Waypoint Model

Table 3.15. Degree of total updates profile

mobility Model	***Mean***	***Median***	***Standard Deviation***	***Skewness***
Random Walk	0.43	0.000	3.14	10.53
Random Waypoint	0.26	0.00	1.41	7.81
Gauss Markov	1.1	0.00	10.91	11.82
D-GM	0.92	0.00	9.73	13.83

3.4.5 Spatial Node Distribution

Visualizing the density of nodes through simulation gives a clearer picture of how nodes are distributed over the simulation period. Moreover, it also presents a different node

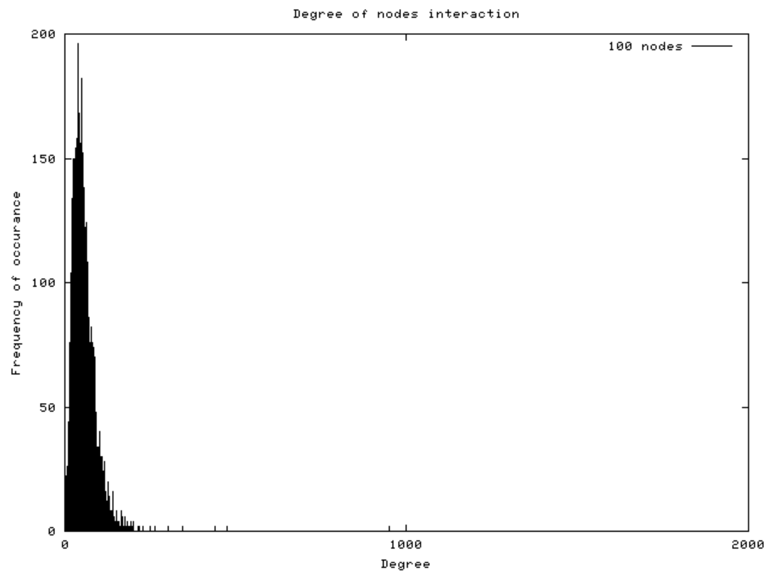


Figure 3.23. Degree of total updates profile using Gauss Markov model

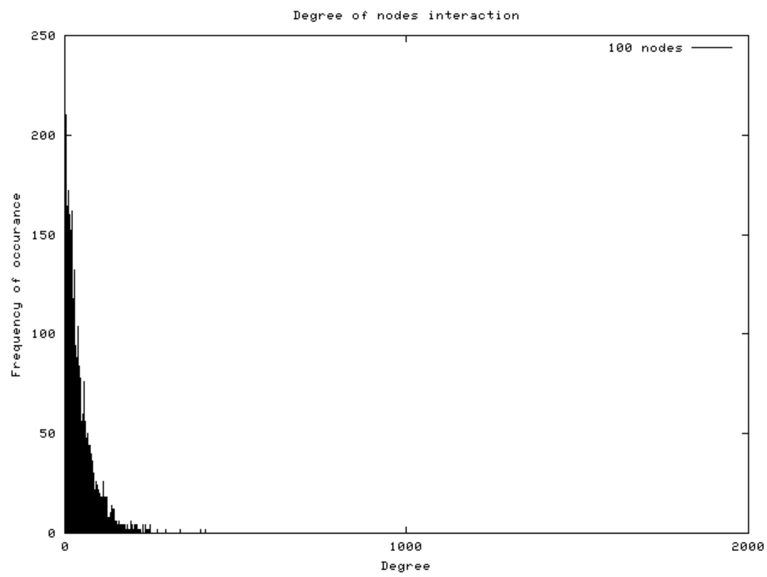


Figure 3.24. Degree of total updates profile using D-GM model

distribution pattern of different mobility models. Figure 3.25, 3.26, 3.27 and 3.28 show the spatial node distribution considering a 9000 step (15 minutes in real time) simulation. We divided the entire simulation area (500m x 500m) into square cells of size 20m x 20m. For each step, the number of nodes that reside in particular cell is counted and added to the respective cells total. The statistics of each figure is presented in Table 3.16. The mean value measure the average of node frequency visiting a cell throughout the simulation.

From Figures 3.25 and 3.26, we can observe both Random Walk and Random Waypoint

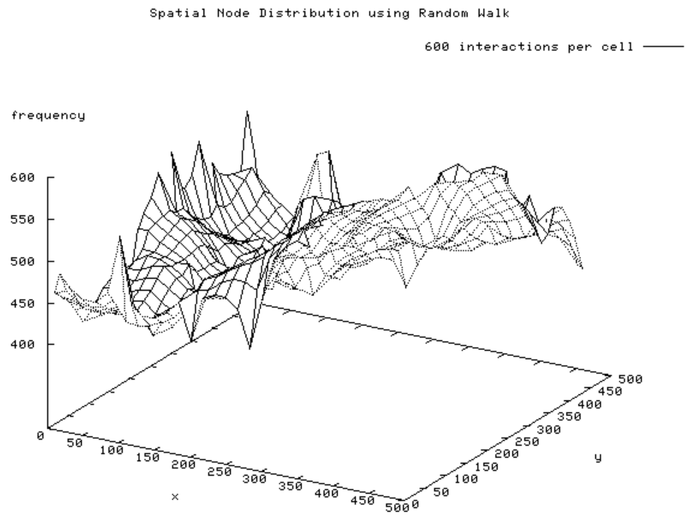


Figure 3.25. Average of spatial node distribution for 100 nodes using Random Walk

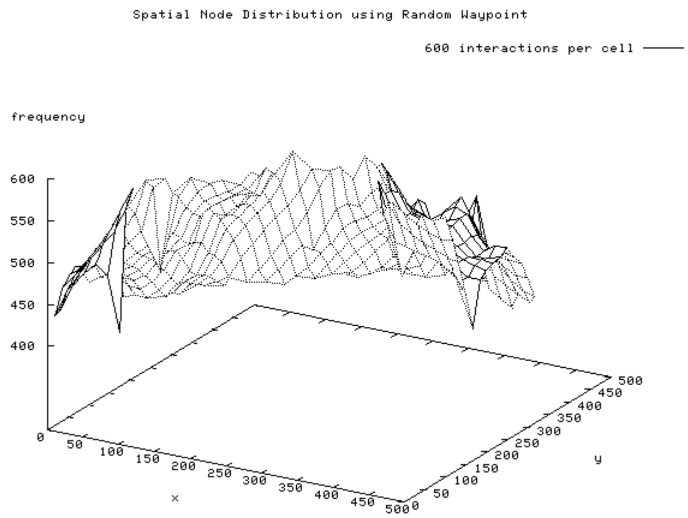


Figure 3.26. Average of spatial node distribution for 100 nodes using Random Waypoint

Table 3.16. Average of Node Distributions using different mobility models with 100 nodes and 9000 simulation steps

	Random Walk	***Random Waypoint***	***Gauss Markov***	***D-GM***
Mean	510.84	516.75	969.40	619.05
Standard Deviation	471.85	33.21	2759.14	1508.74
Median	251	14.62	4	251
Mode	127	14.69	0	0
Minimum	0	0	0	0
Maximum	7558	16888	62950	31649

have a random pattern of node distributions. This is because the selection of the next movement is totally based on a random selection. For the Gauss Markov model, we can

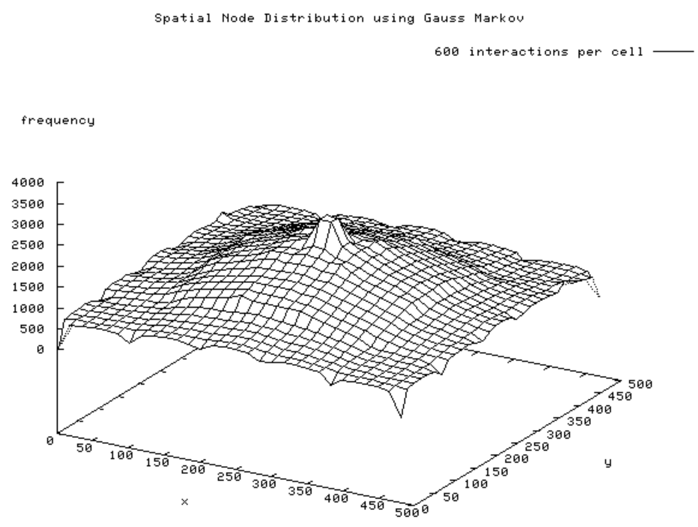


Figure 3.27. Average of spatial node distribution for 100 nodes using Gauss Markov

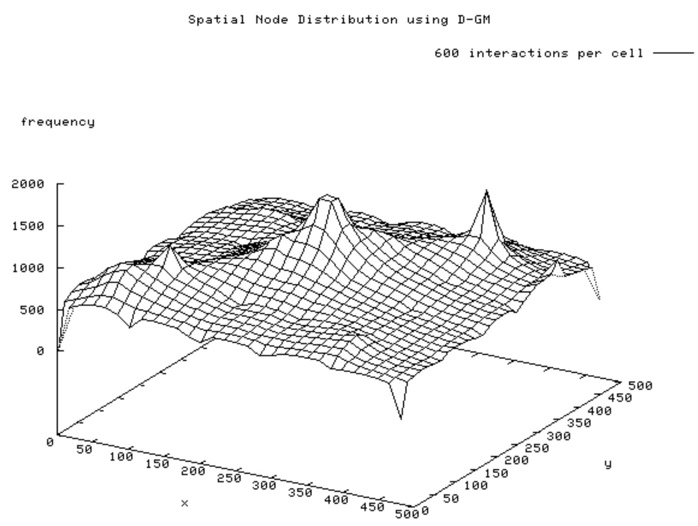


Figure 3.28. Average of spatial node distribution for 100 nodes using D-GM

see from Figure 3.27 the nodes are concentrated in the middle of the simulation space. This is because after a node visits the simulation boundary, a node changes its direction towards the center. The D-GM model has different node concentration distributions (as seen in Figure 3.28) because nodes are moving towards the destination list.

From the statistics in table 3.16, we observe that the Gauss Markov model has a higher maximum value of node visiting a particular cells. This is indicates that more frequent nodes are visiting the same cell (place). This is good for the information dissemination purposes. The second highest maximum value is D-GM model. This is because the D-GM model has set the number of destinations that the nodes might be visiting, so the chance of nodes visiting the same place are very high.

3.5 Conclusion

In this chapter, we have considered the effects of different mobility models with different levels of density. Four mobility models have been evaluated, Random Walk, Random Waypoint Gauss Markov and D-GM. These mobility models have different effect on data dissemination behaviour.

The Gauss Markov outperforms others mobility models in terms of the information dissemination performance. This is because it creates more opportunity of nodes to interact with different nodes. Moreover the movement procedure that forces the direction of nodes to move to the center when they are close to the simulation boundary, ensuring that every nodes has a chance to discover information from the information source.

The Random Walk mobility model disseminates data more quicker than Random Waypoint model. The random assignment of speeds and directions which repeatedly change every 30 seconds creates more opportunities for a node to discover different nodes. This leads to efficient data dissemination. In contrast, even though Random Waypoint has the same characteristics as Random Walk, it has a low data dissemination performance because of the pause attributes. The pause attribute limits a node's opportunity to interact with different nodes as it has to stay in a particular location for a period of time (30 seconds).

In terms of the artifact age, the Gauss Markov has a high number of low age artifact compared to the other models. This is because the nodes change their general direction towards the center when they are appear close to the edge of the simulation boundary. Hence, most of the nodes will have a chance to interact with the information source (that situated at the center of the simulation plane) which enable the nodes to have a fresh artifact from the information source.

In terms of the number of updates, the four models show a positive reaction as the density of node is increased. Gauss Markov generates a high number of updates at the early stage. This is because the nodes are force to move towards the center after visiting the simulation boundary. This creates a high opportunity for nodes to be frequently updated by the nodes that already have a fresh from the infostation which located in the middle of the simulation.

Because this works is based on a computer simulation, the results here are not fully realistic for all types of human mobility. However, it is appropriate approach to understand the behaviour of data dissemination for different possible scenarios because it has the flexibility on examining different effects of different settings in effective way (i.e. fast,less cost). Moreover, there is no unique way to model the human's mobility.

In conclusion, the data dissemination behaviour is sensitive to the mobility model and the node density. These two factors are important to be included when investigating the information dissemination behaviour in opportunistic networks. In this chapter, only single data dissemination protocol is used (i.e flooding). This protocol only pushes information blindly to its direct neighbours. We believe that using different ways of information exchange protocol will affect the information dissemination behaviours or performance. So, in the next chapter (Chapter 4) we investigate further the effect of different ways of acquiring information (i.e push and query) on information dissemination behaviour. Because of the findings in this chapter we will continue to use all four mobility models in the thesis.

Chapter 4

BASELINE APPROACHES FOR DISSEMINATION

4.1 Introduction

The purpose of this chapter is to introduce four new baseline data dissemination techniques in mobile peer to peer communication. We call these are Pure Push, Greedy, L-Push and Spray and Relay. These techniques are simple with limited intelligence. Simplicity is important because this well-suite the time and resource constrained environment. In this chapter these techniques are formed to investigate the effect of data dissemination using different ways of spreading information. We purposely designed these dissemination techniques to focus on the Query and Push processes, which are the key components that need to be controlled to spread information with less overheads in mobile peer to peer networks. One of our techniques (Spray and Relay) is a modified version of the Spray and Wait technique [43]. Our technique modified the Spray and Wait technique in terms of Push and Query, quota and its usage. These will underpin future comparison in our study.

4.2 Dissemination Techniques

Push and Query processes are functions that can occur at a node when two nodes are in direct transmission range. It is assumed that information is homogenous and equally wanted by all nodes. What we are interested is to find out how simple Push and Query processes affect information spreading. A Query is a message that is used to ask for new (or updated) information. It is a simple way for a node to quickly know its peers preference at a particular point of time. However, it needs a proper mechanism to control the query to avoid a *query storm*, where saturation of Query messages used excessive resources and bandwidth. Push is the process of forwarding information to another node. It can occur directly or indirectly in response to a query. A push also can be issued regardless to response to any query. Similarly to a query, Push has to be carefully used to avoid duplication of forwarding content to the same nodes. Therefore the dissemination techniques introduced in this chapter are designed to understand and later (in chapter 5) to utilize the advantages of Push and Query process to disseminate or spread information effectively in an opportunistic network. Table 4.1 summarize the four dissemination techniques introduced in this chapter.

Definition 1. (Query) Let a and b be nodes and assume a and b are in range and a does not have information. Query is a message sent from a in which it requests information from b.

Definition 2. (Push) Assume that a and b are nodes that are in range. A push from a to b is when node a forwards information to b.

Definition 3. (Query Quota) Let a be a node. Query Quota is the number of queries that a is allowed to issue. The Query Quota is decreased by one for every query issued by node a.

Definition 4. (Renew interval time) Let t be a time step and a is a node. Assume that a has just renewed its query quota at time t_i. a can only renew its query quota at time step $t_i + t_j$, where t_j time is called *renew interval time.* At every renew interval, all nodes have renew their query quota .

Table 4.1. Data dissemination techniques attribute (Push and Query)

Dissemination techniques	Push				Query	
	No Quota	With Renew Quota	With Relay quota	Required query to execute push	No quota	With quota
Pure Push	✓	×	×	×	×	×
Greedy	✓	✓	×	✓	×	✓
L-Push	×	✓	×	✓	✓	×
Spray-Relay	×	×	✓	×	×	×

From Table 4.1, there are number of ways in which different combinations of Query and Push can be implemented. Push can be implemented in four different ways.

1. **Push with no quota** - This option allows a node to push information to its current peer at any time. It can push information freely.

2. **Push with renew quota** - This type of Push limits the number of times information can be pushed within the particular time interval (definition 4). The Push process only executed when a node has a quota to push.

3. **Push with relay quota** - The size of push quota (L) is decided at the beginning of the simulation. This quota is not renewable as in Push with renew quota. Remaining quota is also transferable to other nodes when information is pushed. For example, in the Spray and Wait [43] dissemination approach, half of push quota of a source node is transferred to a relay node.

4. **Push when there is a query** - The Push function will be executed when a node received a query.

As for the Push mechanism, Query also can be implemented in two different ways, without quota and with quota. Query without quota permits a node to send a query to its encountered peer without limit. The Query with quota has a renewable quota within the particular time interval. The Query is also used based on the basis of first come first serve. The following section explains the detail of how the dissemination approaches work. The term *current peer* used in the following algorithms refers to the node that is involved in the interaction. The algorithms are executed whenever there is an interaction between nodes.

4.2.1 Pure push Approach

Pure Push is a technique that forwards information whenever it is possible without considering the status (has information or not) of its current peer. Nodes with information are the nodes that are pushing information and nodes without information will be listening and wait to receive information. Eventually, all nodes will tend towards being infected (have information) and active to forward information. At this stage, this *information has been flooded* across the network which causes high push overheads (i.e. possible message duplication). Algorithm 2 shows the detail of how this technique works.

Algorithm 2 Pure Push (Flood)

```
1: if the node has an artifact then
2:     Push the artifact to the current peer
3: end if
```

4.2.2 Greedy Approach

The Greedy technique focuses on query attributes. A node has to receive a query before pushing information. When a node receives a query from its current peer, it will respond (pushing information) if the node has information that matches the query. Otherwise, the node ignores the query. In addition, the query also has a quota which limits the number of queries that a node can issue at the particular time interval. This quota called *query quota*: it is renewable after a certain time interval. The definition of query quota is given in Definition 3. The duration of the renew time interval is between the beginning of a renew quota and the next renew quota. The definition of renew time interval is defined in Definition 4. The detail of how the Greedy technique works is shown in Algorithm 3.

Algorithm 3 Greedy

```
if the node has an artifact then
    if the node received a Query then
        Push artifact to the current peer
    end if
else
    if the node has a query quota then
        Send query to the current peer
        queryquota - 1
        update queryqouta
    end if
end if
```

In terms of dissemination, the Greedy approach has less overheads compared to the Pure Push approach. This is because the information is forwarded from the source to its destination only when there is a query received by the source nodes. The query issued depends on the query quota of a node. The more the query quota is, the higher the chance that a node will discover information. On the other hand, increasing the query quota will also increase the query overheads. So, the renew interval quota is an important parameter to help to understand the performance of data dissemination. The shorter the renew interval quota the more chance for the node to discover information. This is because nodes will only have a short time to regain its query quota.

4.2.3 L-Push

L-push is a kind of Greedy technique which has query and push capability. However L-Push has no quota on the query (contrast with Greedy) but it has a quota on the Push. Through this technique, we can examine the effect of the Push on the data dissemination performance. The Push quota works in a similar way to the Query quota in the Greedy technique, where it is renewable. Algorithm 4 shows the process of how this technique performs.

L-Push has advantages in terms of discovering information because it does not have a quota on the query. This helps nodes easily to discover information in a short while. Since the query is unlimited, the Query message overheads and message duplication will be higher as Push information in Pure Push. The data dissemination rate is dependent on the Push quota because even if a node receives a query, no information will be forwarded

Algorithm 4 L-Push

if *the node has an artifact* **then**
  No Query issued
  if *the node received a Query* **then**
    if *the node has push quota* **then**
      Push the artifact to the *current peer*
      push quota - 1
      update the *push quota*
    end if
  end if
else
  Send Query to the current peer
end if

unless a node has a push quota to push information. Therefore the Push quota is a strong factor that influence the data dissemination performance of L-Push technique.

4.2.4 Spray and Relay

Spray and Relay is a modification of the Spray and Wait technique [43]. The Spray and Relay attributes are similar to L-Push. However, instead of the Push quota being renewed at every time interval, the quota in Spray and Relay is relayed to nodes with the information that being pushed. The quota that is passed to the next node is a half of the nodes' quota. When a node's quota becomes zero, a node stops forwarding information. The following are our assumptions on the Spray and Relay based on our perspective:

1. L is quota for forwarding (i.e. pushing).
2. Information source only forwards quota to a node when it first makes contact with it. After that, the information source will not issue any information.
3. Nodes can forward a proportion of quota when pushing.
4. A node stops forwarding information when $L = 0$.
5. The quota is not renewable and is only issued once by the information source.

The Spray and Wait technique that is discussed in [43] has the ambiguity of how a node knows that its peer already has information. So, we introduce a Query concept to make this issue clear. Further more, we also explicitly used ideas of Push, Query and Quota to resolve this problem. Algorithm 5 shows the Spray and Relay process in detail.

Algorithm 5 Spray and Relay

```
if the node has an artifact then
    if the node receives a query then
        if L >= 1 then
            if L = 1 then
                Push artifact with quota L = 1 to the current peer
                Set current Node quota L = 0
            else
                Push the artifact with quota L = L/2 to the current peer
                Set Node quota L = L/2
            end if
        end if
    end if
else
    Send a query to the current peer
end if
```

Using the Spray and Relay approach, we can evaluate exactly the effect of Push quota on the data dissemination performance. Note that the Push quota is not renewable but transferable from one node to other node and therefore the amount of pushing can be controlled. In the case $L/2$ is odd, the value of L will be rounded down. However, this approach suffers from a large number of queries because nodes without information will always issue a Query. In addition, determining the best Push quota is a challenge to use this approach and this affects the dissemination performance.

4.3 Key Performance Indicator (KPI)

The KPI in this chapter is focused on data dissemination performance and overheads. These KPIs are used to identify which techniques perform better in different scenarios.

4.3.1 Average number of nodes with an artifact at time t

This KPI measures how quickly the information is spread in a given test problem. The information (artifact) is time independent which means that once a node receives an artifact, that artifact will be valid though the simulation time. At each time step, the number of nodes that have an artifact is counted and recorded. This step is performed for every trial. At the end of the number of trials, the numbers of nodes that have an artifact at every time step are summated and then divided by the number of trials. Equation 4.3.1 shows the average number of nodes that have an artifact at time t is measured.

$$AN_t = \frac{\sum_{n=1}^{TR} AN_{n,t}}{TR} \tag{4.3.1}$$

where TR is the number of trials, AN_t is the average number of nodes that have the artifact at time step t and $AN_{t,n}$ is the number of nodes that have an artifact at time t in trial n.

4.3.2 Average of Coverage of an Artifact

This KPI measures the average number of nodes that discover an artifact over TR simulation trials. For every trial, the number of nodes that have an artifact is counted and recorded for each time step of the trial. These values are totaled per trial and averaged over all trials. Equation 4.3.2 shows the average of coverage of an artifact is measured.

$$AC = \frac{\sum_{n=1}^{TR} \sum_{t=0}^{ST} AN_{n,t}}{TR} \tag{4.3.2}$$

where TR is the number of trials, ST is the length of the simulation.

4.3.3 Average Push Overhead

Push overhead measures how many pushes were involved in the process of disseminating information to all nodes. The greater the number of pushes involved, the more energy will be consumed within the system as a whole. Therefore, this KPI helps us to identify the performance concerned with the number of pushes for different scenarios. To measure this KPI, we recorded every interaction in every time step across the simulation. Then we average the push with the number of simulation trials. The following shows how the average of overhead interactions is calculated in our simulation.

$$AP_t = \frac{\sum_{n=1}^{TR} P_{t,n}}{TR} \tag{4.3.3}$$

where TR is the number of trials, AP_t is the average number of Pushes at time step t and $P_{t,n}$ is the number of Pushes at time t with n trial.

4.3.4 Average Query Overhead

This KPI measures the number of Query overheads involved in the dissemination techniques. The higher the query overheads, the more energy will be consumed. Therefore, using this KPI, we are able to identify the best Query quota setting for different scenarios (i.e. with low average Query overheads). We recorded every Query generated in every time step. Then, we average the number of Query used across the each of simulation trials. The following Equation 4.3.4 is used to calculate the average of Query overheads.

$$AQ_t = \frac{\sum_{n=1}^{TR} Q_{t,n}}{TR} \tag{4.3.4}$$

where TR is the number of trials, AQ_t is the average number of queries at time step t and $Q_{t,n}$ is the number of Query at time t with trial n.

4.4 Experimentation

The organization of the experiments in this section is based on the dissemination techniques (i.e. Pure Push, Greedy, L-Push and Spray and Relay) that are introduced in this chapter. The main purpose of the experiments is to examine the best combination of parameters for each approach in order to achieve the best individual performance in spreading information while minimizing the usage of resources (i.e. number of queries and pushes). The following sections are laid out with different parameters (according to the approach attributes) to investigate the highest performance combinations. The duration of simulation for every experiment section is 9000 simulation time steps (15 minutes in real time) and each simulation is repeated 50 times (i.e. 50 trials) and at every trial we using different seed to have unbias experimental results. The same seed are used for all mobility models. The test scenario considers 100 nodes randomly place in 500 square meters. The information source is placed in the middle of the simulation plan at coordinates x: 250 and y: 250. We use the four different types of mobility models: Random Walk, Random

Waypoint, Gauss Markov and D-GM model.

Throughout the results, we *accumulate cost of overhead* (in units) to measure the overheads for each technique. The cost consists of two elements, Push overhead and Query overhead. We assume Push and Query rate as follows to calculate the cost.

- *1 push = 1 unit used of resources*
- *1 query = 0.5 units used of resources*

we note that other alternatives could be used and this assumes relatively small payloads are used.

Up to the best of our knowledge, the approach that we introduce in this chapter have not been addressed by other researchers. Therefore, there is no best parameters setting to conduct experiments for each approach, we varied our push and query quota between maximum and minimum levels. We choose the maximum level to be 100 because the number of mobile nodes in our experiment is 100. So, if in every time step nodes meet different nodes in the simulation, all nodes will discover an artifact within 100 time steps. This would give the fastest artifact discovery when several nodes are pushing at the same time. The minimum level of quota is 2. We choose 2 as a minimum level of quota because we assume that the first quota might be not effective, so by giving the second chance quota will increase the chances of discovering and pushing information. For the interval time between nodes renewing their quota, we set the range between 5 and 200 seconds. This is because we want to see how the small interval values influence the information dissemination. The maximum level of interval time is 200 for Pure Push protocol experiment. The reason why we choose 200 time interval is because with 200 interval time we can make sure less than 45 times each node has opportunity to renew its push or query quota. This value is not the best but we use it as to make our experiments more manageable and can be changed in the future.

4.4.1 Results-Pure Push

Pure Push is an approach that pushes information whenever it is possible without any limit. No Query is included in this approach. This approach sends information to a single direct peer once the channel between pair is set up. This is slightly different to broadcasting where all peers that are in range receive the message.

The main purpose of this experiment is to examine the effect of Push mechanism on data dissemination performance. The results of this experiment is used as a benchmark to evaluate the performance of the other approaches (Greedy, L-Push and Spray and Relay). The results of the experiments are tabulated in Table 4.2.

Table 4.2. Statistics for Pure Push Approach.

Mobility Model	**Average of coverage at each time step**	**Standard Error**	**Standard Deviation**	**Push Cost Overhead (units)**
Random Walk	88.489	0.268	25.449	417978.48
Random Waypoint	81.495	0.307	29.158	391267.04
Gauss Markov	98.870	0.096	9.145	548023.12
D-GM	98.329	0.098	9.321	212320.1

From Figure 4.1 we can observe that:

- Gauss Markov outperforms other mobility models in spreading information. This is due to the fact that all nodes that move close to the boundary of the simulation, change direction towards the center. Therefore, the chances of nodes meeting (at the center) and discovering new information from other nodes are very high. Random Waypoint has the slowest performance in information dissemination. This is because nodes are forced to stop at certain places for a period of time before moving to another place. Therefore, this limits the opportunity of nodes to discover more nodes which potentially increases the information dissemination performance.

- In terms of performance information availability, Gauss Markov model achieved complete dissemination in less than 50 seconds. A complete dissemination stage is where all nodes have information. This is followed by Random Walk, D-GM and Random Waypoint which respectively reach the completed information dissemination in less than 400, 500 and 600 seconds. The high frequency of node interactions in the Gauss Markov model helps the information spread quickly. Further more, the information source that is located in the middle of the simulation is also contributes to the acceleration of the information dissemination because the Gauss Markov model forces all the nodes to move to the center of the simulation area when the nodes are closed to any the simulation boundaries.

- From Table 4.2, D-GM has the lowest Push overhead costs. This is because some nodes are not involved in pushing artifacts as they are not in range of other nodes. This reduces duplicate messages in the D-GM model compared to other models. Comparing the GM and D-GM model, they have only 0.541 difference in the average level of coverage. This shows that the D-GM is able to disseminate information with lower overhead cost with a reasonable performance relative to the Gauss Markov model.

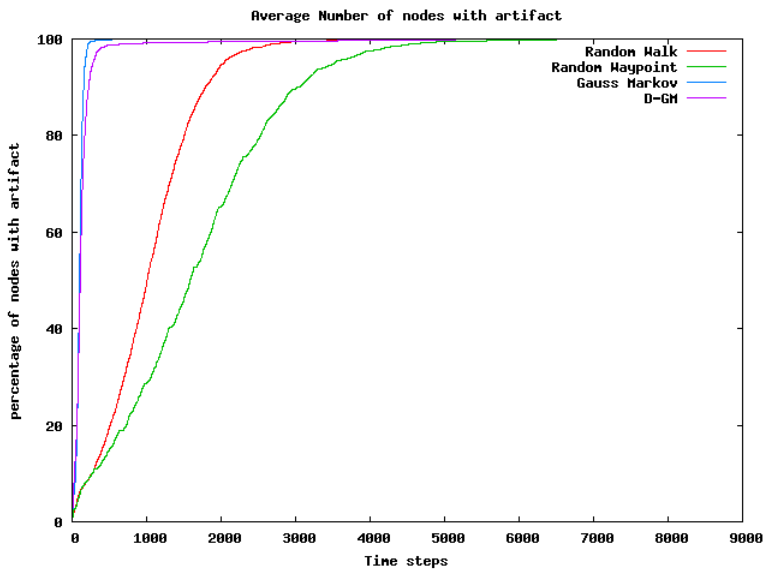


Figure 4.1. Average number of nodes with artifact using Pure Push approach

4.4.2 Result-Greedy

In the Greedy approach, nodes are limited by quota in sending a Query but unlimited in pushing when they receive a query. There are two parameters that can vary the performance of this approach. They are *Query quota* and *Interval time* which controls the number of query and the renewal of query quota respectively. The following subsections specifically examine the effects of these parameters on the information spreading performance.

4.4.2.1 The effect of Query quota on information spreading.

The purpose of this experiment is to find out the effect of Query quota in Greedy approach on information spreading performance. We set the Query quota per node at various settings between the interval [2-100] and renewed for every 200 time steps (time intervals). Different mobility model also used in this experiments. The results are tabulated in Table 4.3. Note that the pushes in Table 4.3 does not include pushes from information source.

Table 4.3. Statistics for different Query quota with different Mobility Model using Greedy Approach.

Mobility Model	**Query Quota**	**Average artifact coverage**	**Standard Error**	**Standard Deviation**	Accumulative Cost of overheads	
					Push	Query
Random Walk	2	83.787	0.324	30.773	95.16	648.14
	5	84.433	0.317	30.105	95.54	2961.16
	7	84.854	0.313	29.682	95.8	3975.86
	9	85.247	0.309	29.205	96.02	4930.4
	40	87.623	0.277	26.312	97.32	16482.14
	70	88.377	0.269	25.586	97.56	23171.26
	100	88.489	0.2683	25.449	97.6	25092.76
Random Waypoint	2	72.842	0.362	34.211	95.5	1984.38
	5	74.525	0.353	33.528	95.72	4499.32
	7	75.193	0.349	33.133	95.86	6058.14
	9	75.902	0.346	32.789	95.98	7507.4
	40	80.208	0.318	330.123	97.5	24937.94
	70	81.297	0.309	29.292	97.6	35962.4
	100	81.494	0.307	29.159	97.6	39399.92
Gauss Markov	2	96.751	0.160	15.168	78.26	399.42
	5	97.408	0.147	13.938	82.76	751.78
	7	97.581	0.144	13.671	84.0	959.58
	9	97.724	0.142	13.440	85.12	1146.22
	40	98.672	0.100	9.530	95.92	2807.8
	70	98.868	0.096	9.146	97.2	2957.66
	100	98.870	0.096	9.145	97.2	2961.74
Directed-Gauss Markov(D-GM)	2	95.548	0.149	14.159	70.2	371.24
	5	96.088	0.142	13.437	76.06	734.56
	7	96.379	0.139	13.155	78.14	947.02
	9	96.656	0.136	12.931	79.62	1141.54
	40	96.143	0.142	13.483	95.26	2780.32
	70	96.131	0.140	13.313	96.52	2983.06
	100	96.132	0.140	13.307	96.54	2988.44

Figures 4.2, 4.3, 4.4 and 4.5 show the effect of different Quota on information spreading. From the figures we observed that:

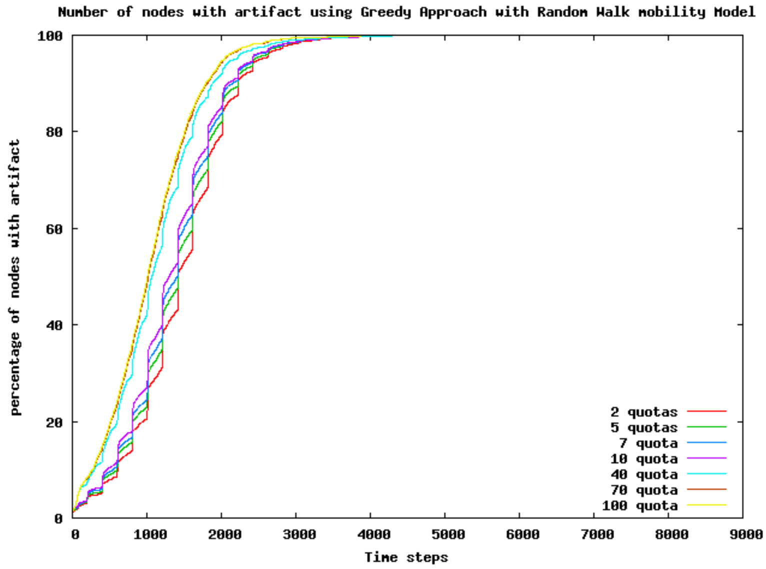


Figure 4.2. Average number of nodes with artifact using Greedy Approach with Random Walk while varying query quota

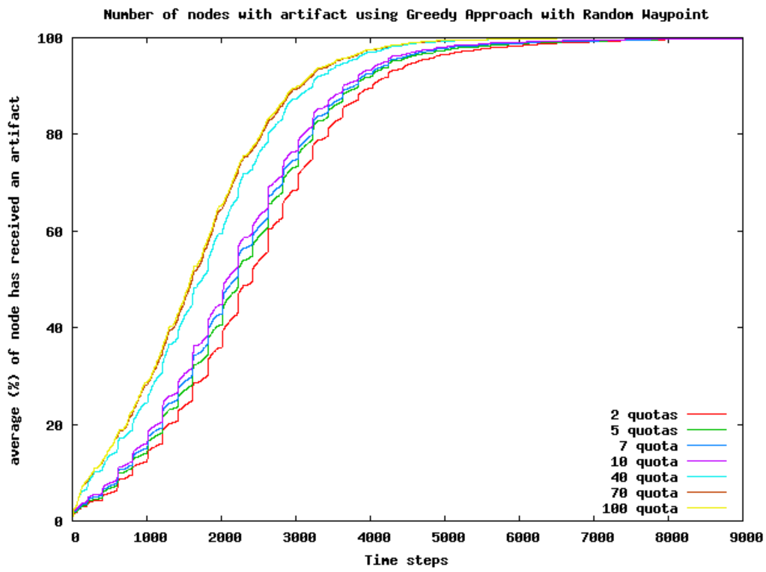


Figure 4.3. Average number of nodes with artifact using Greedy Approach with Random Waypoint while varying query quota

- For every mobility model, when the Query quota is increased, the number of nodes with information is also increased. This is due to the fact that increasing Query quota creates more opportunity for nodes to discover new information through sending multiple queries. Based on Figure 4.2 and 4.3, there is a small change in the number of nodes with information when the Query quota is set between 2 to 9. However when the Query quota is set into 40, it creates a large gap. This is because the Query

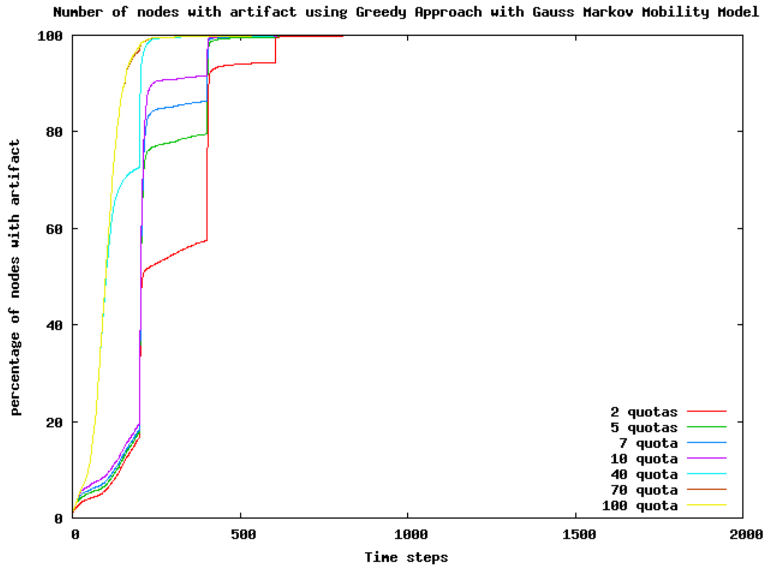


Figure 4.4. Average number of nodes with artifact using Greedy Approach with Gauss Markov while varying query quota

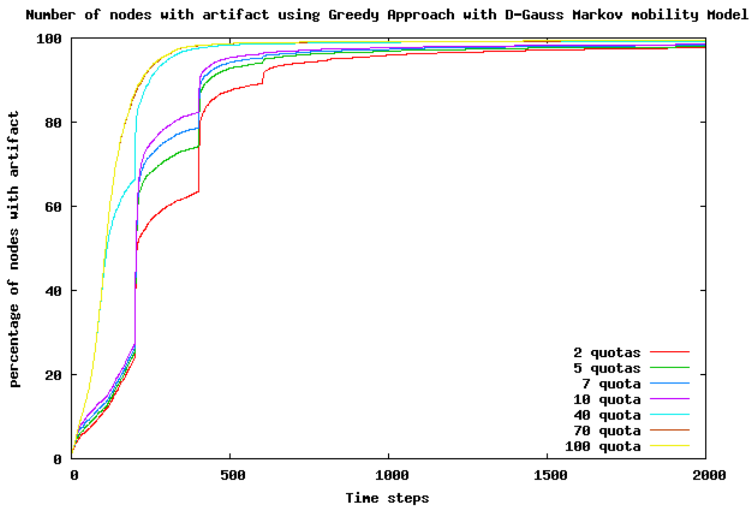


Figure 4.5. Average number of nodes with artifact using Greedy Approach with Directed Gauss Markov while varying query quota

quota at 40, has a massive number of queries issued compared to the Query quota between 2 and 9. This is shown in Table 4.3 in the Query column (last column). For example, looking at the Random Walk row in Table 4.3, the query cost when the Query quota is equal to 9 is 4930.4 and the query cost is 16482.14 when the Query quota is equal to 40. Note that there is a large difference in the number of queries issued between them and this causes the big increase in artifact dissemination as seen in Figure 4.2. This is the same case for Random Waypoint, as shown in Figure 4.3. For the Gauss Markov and D-GM, the change of Query quota is not showing any significant differences. However, increasing the Query quota is still increases the number of nodes with artifact. Note that Figures 4.4 and 4.5 are scaled across 2000 steps rather than 9000 steps.

- In terms of information availability, Guass Markov spreads information very fast at an early stage and reaches its optimum in less than 50 seconds as shown in Figure 4.5. D-GM has the same pattern as Guass Markov but reaches its optimum at nearly 100 seconds. Random Walk and Random Waypoint consistently increase the dissemination of information.

- In terms of overhead costs, Random Waypoint is the highest. Followed by Random Walk, D-GM and Gauss Markov. From Table 4.3, even though the Random Waypoint generates more queries but it still has the lowest average of artifact coverage. This tells us that there is a high duplication (unnecessary queries issued) in Random walk because Guass Markov has the lowest overhead cost is able to reach a higher number of nodes with an artifact which up to 98.87 percent.

4.4.2.2 The effect of Interval Time on information spreading.

Interval time is the duration where the node will renew its Query quota. This parameter is important to be examined because it influences the ability performance of a node to discover information. We use 100 nodes with 50 time trials and using four different mobility models for this experiment. In addition, we varied the interval time so that we can identify the best performance interval time to spread information using the Greedy approach. We choose Query quota = 5 for every experiment because we found from the experiments in 4.4.2.1 that is gives a reasonable artifact coverage with a small of overheads.

Figures 4.6, 4.7, 4.8 and 4.9 show the effect of interval time on information spreading. From the figures we observed that:

- Gauss Mobility model outperforms other models in terms of information spreading time. Based on Figure 4.8, it takes less than 250 seconds to disseminate information

Table 4.4. Statistics for different Interval Time with different Mobility Models using Greedy Approach.

Mobility Model	**Interval Time**	**Average artifact coverage**	***Standard Error**	**Standard Deviation**	Accumulative Cost of overheads	
					Push	Query
Random Walk	5	87.42	0.283	26.841	97.16	7312.44
	10	86.068	0.299	28.379	96.62	4604.66
	20	84.433	0.317	30.105	95.54	2961.16
	30	82.909	0.329	31.174	94.86	2350.3
	60	78.314	0.357	33.855	92.56	1648.64
	90	72.251	0.380	36.036	90.28	1464.68
	120	67.354	0.386	36.657	88.94	1322.98
Random Waypoint	5	80.236	0.317	30.045	97.26	10991.86
	10	78.10132	0.331	31.423	96.84	6816.14
	20	74.525	0.353	33.528	95.72	4499.32
	30	71.664	0.363	34.392	95.32	3605.14
	60	66.504	0.375	35.564	94.12	2397.16
	90	60.655	0.379	35.953	91.98	1970.44
	120	55.407	0.374	35.479	89.64	1726.44
Gauss Markov	5	98.457	0.114	10.835	95.22	1406.94
	10	98.088	0.127	12.058	90.66	994.26
	20	97.408	0.147	13.938	82.76	751.78
	30	96.484	0.168	15.974	76.28	705.82
	60	94.273	0.207	19.614	63.06	626.16
	90	92.514	0.227	21.553	51.92	584.46
	120	91.043	0.239	22.638	44.34	556.64
Directed Gauss Markov	5	97.794	0.113	10.733	93.54	1366.82
	10	97.318	0.123	11.708	87.1	965.22
	20	96.088	0.142	13.437	76.06	734.56
	30	94.799	0.156	14.757	65.56	650.68
	60	91.022	0.178	16.923	51.16	586.68
	90	91.022	0.1783	16.924	51.16	586.68
	120	87.091	0.192	18.173	36.58	540.12

to all nodes, whereas Random Walk (Figure 4.6) and Random Waypoint (Figure 4.7) required more than 900 seconds to complete the dissemination. This is indicates that Gauss Markov is less affected when the interval time is varied compared to other mobility models.

- In term of performance, when the interval time is increased, the information spreading is decelerated. This is due to the fact that a short interval time increase the capability of nodes to send a query as it can renew it Query quota more frequently. There-

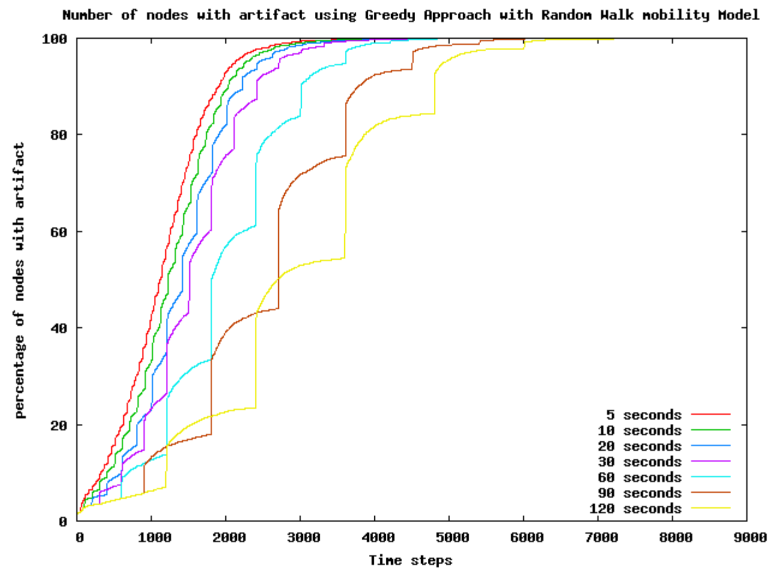


Figure 4.6. Average number of nodes with an artifact using different Interval times and the Greedy Approach with Random Walk

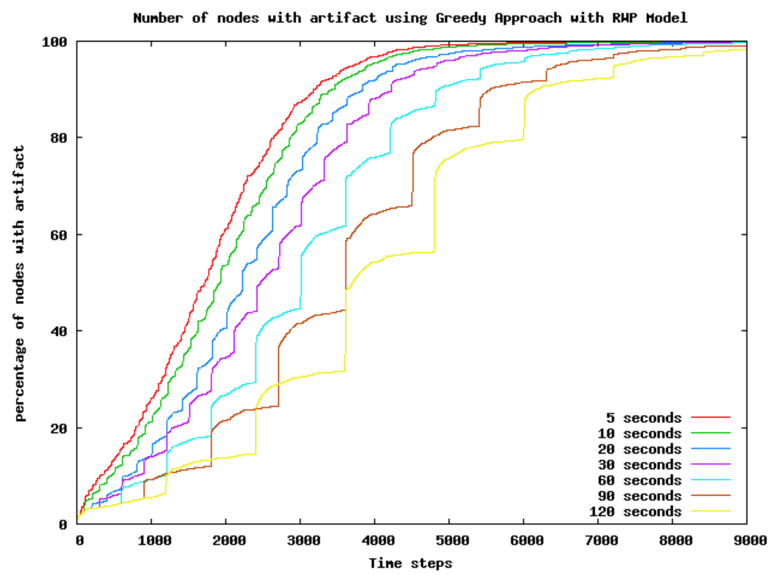


Figure 4.7. Average number of nodes with an artifact using different interval times and the Greedy Approach with Random Waypoint

fore, a node has a large opportunity to discover information very quickly. Whereas, if the interval time is big, it holds the opportunity of nodes to send more query as it needs to wait for a long period of time until it regain its Query quota.

- From Table 4.4, the accumulative overhead costs for all mobility models are decreased when the interval time is increased. This is because Query and Push is constrained by the Interval time. For example, in Gauss Markov model, when the interval time

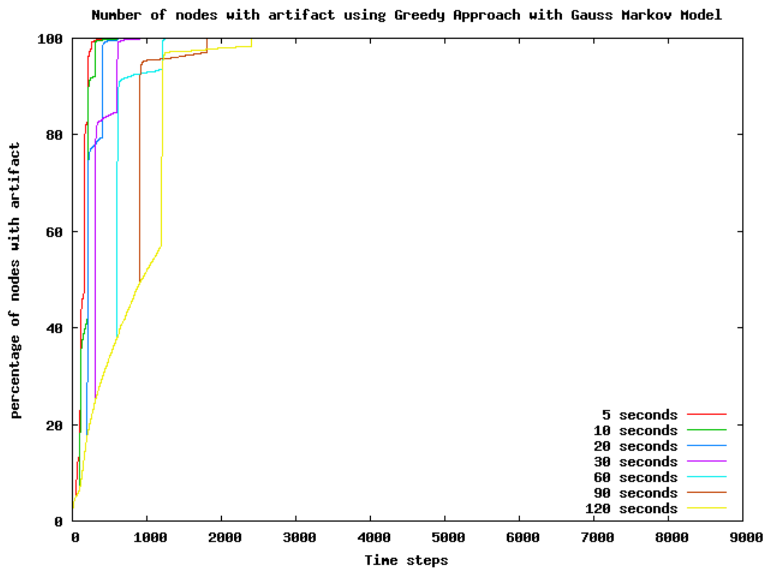


Figure 4.8. Average number of nodes with an artifact using different interval times and the Greedy Approach with Gauss Markov

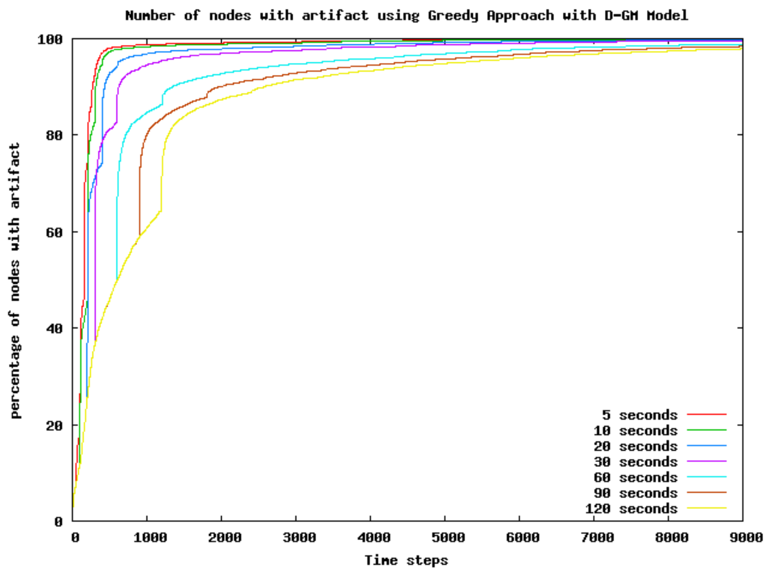


Figure 4.9. Average number of nodes with an artifact using different interval times and the Greedy Approach with Directed Gauss Markov

is set at 5 seconds the number of queries issued is 1406.94 units and when the Interval time is increased up to 30 seconds, the number of query issue is dropped to 705.82 units. This tells us that controlling the number of query will slow down the information dissemination as it limits the nodes interactions. Further more, we learned that controlling the frequency of Push and Query can reduce the overhead costs.

From the results presented in Section 4.4.2.2, we can observe the best combination settings for Query quota and Interval time to have a high performance using the Greedy approach. We select the *average artifact coverage* and *accumulative cost of overheads* from Table 4.4 and 4.3 to identify the best combination. The selection is focused on Gauss Markov and D-GM mobility models. From Table 4.3, the selection is based on a trade-off between minimizing the query quota and maximizing the performance. **We pick the query quota equal to 5 and the interval time equal to 20 seconds** because this combination has better average artifact coverage performance with a small number of query quota.

4.4.3 Results L-Push

L-Push has a Push quota but has unlimited Query quota. Push process is only triggered when there is a query from other node. There are two attributes that can vary the performance of this approach. They are the size of Push quota and the Interval time. We use experiments to investigate the effect of the different setting on information spreading. The first experiment investigates the effect of Push quota on information spreading. The second experiment is focused on examining the effect of Interval time on information spreading. Table 4.5 shows the range parameter settings for this experiments.

Table 4.5. L-Push experiment Parameter Settings

Parameters	Settings
Mobility Model	Random Walk, Random Waypoint, Gauss Markov and Directed Gauss Markov
Number of nodes	100
Number of trials	50
Push quota	2,5,7, and 9
Interval Time	20 seconds (200 time steps)

4.4.3.1 The effect of Push quota on information spreading

Push is required for information dissemination. However it needs to managed properly, otherwise many unnecessary messages will be in the network. Therefore it is vital to know exactly what is the best Push quota that is needed to spread information effectively. Knowing the best performance Push quota also will reduce the total overhead costs. The results of this experiment are tabulated in Table 4.6.

Table 4.6. Statistics for different Push quota with different Mobility Models using L-Push Approach.

Mobility Model	**Push Quota**	**Average of artifact coverage**	**Standard Error**	**Standard Deviation**	Accumulative Cost of overheads	
					Push	Query
Random Walk	2	88.459	0.269	25.494	97.6	112.74
	5	88.489	0.268	25.449	97.6	97.6
	7	88.489	0.2683	25.449	97.6	97.6
	9	88.489	0.268	25.449	97.6	97.6
	40	88.489	0.268	25.449	97.6	97.6
	70	88.489	0.268	25.449	97.6	97.6
	100	88.489	0.268	25.449	97.6	97.6
Random Waypoint	2	81.493	0.307	29.161	97.6	102.7
	5	81.495	0.307	29.158	97.6	97.6
	7	81.495	0.307	29.158	97.6	97.6
	9	81.495	0.307	29.158	97.6	97.6
	40	81.495	0.307	29.158	97.6	97.6
	70	81.495	0.307	29.158	97.6	97.6
	100	81.495	0.307	29.158	97.6	97.6
Gauss Markov	2	98.749	0.100	9.532	96.98	261.54
	5	98.866	0.0965	9.151	97.2	104.0
	7	98.870	0.096	9.146	97.2	97.38
	9	98.870	0.096	9.145	97.2	97.2
	40	98.870	0.096	9.145	97.2	97.2
	70	98.870	0.096	9.145	97.2	97.2
	100	98.870	0.096	9.145	97.2	97.2
Directed Gauss Markov	2	97.895	0.105	9.976	96.34	378.06
	5	98.285	0.099	9.368	96.54	119.76
	7	98.319	0.098	9.325	96.54	99.52
	9	98.329	0.098	9.321	96.54	97.0
	40	98.329	0.098	9.321	96.54	96.54
	70	98.329	0.099	9.321	96.54	96.54
	100	98.329	0.098	9.321	96.54	96.54

Figures 4.10, 4.11, 4.12 and 4.13 indicate the effects of Push quota on the information spreading. From the figures we can observe that:

- In general, different mobility models have different effects on information spreading. Mobility models determine the possible interactions between nodes. The higher the interaction is the higher the possibility of nodes able to discover new information.

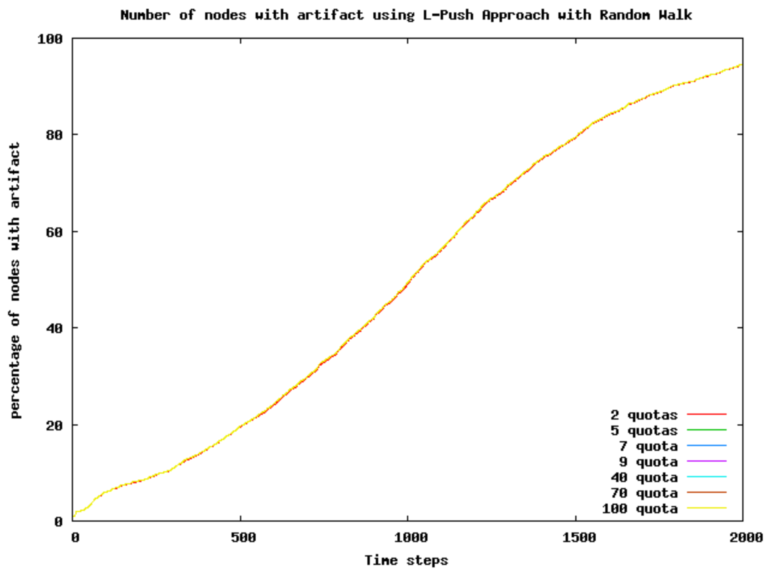


Figure 4.10. Average number of nodes with an artifact in using L-Push Approach with Random Walk while varying push quota

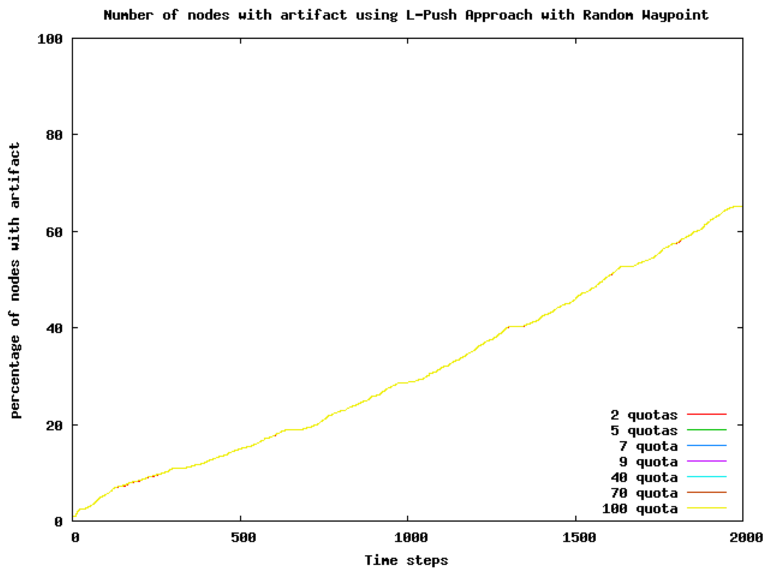


Figure 4.11. Average number of nodes with an artifact using L-Push Approach with Random Waypoint varying push quota

The order of the best mobility model is Gauss Markov, Directed Gauss Markov (D-GM), Random Walk and Random Waypoint.

- From the Figures 4.10 and 4.11, there is no obvious change for Random Walk and Random Waypoint in the number of nodes with information when varying the push quota at different levels. This is because the L-Push approach only uses its push quota when it is necessary. Since the L-Push has unlimited Query quota, a node

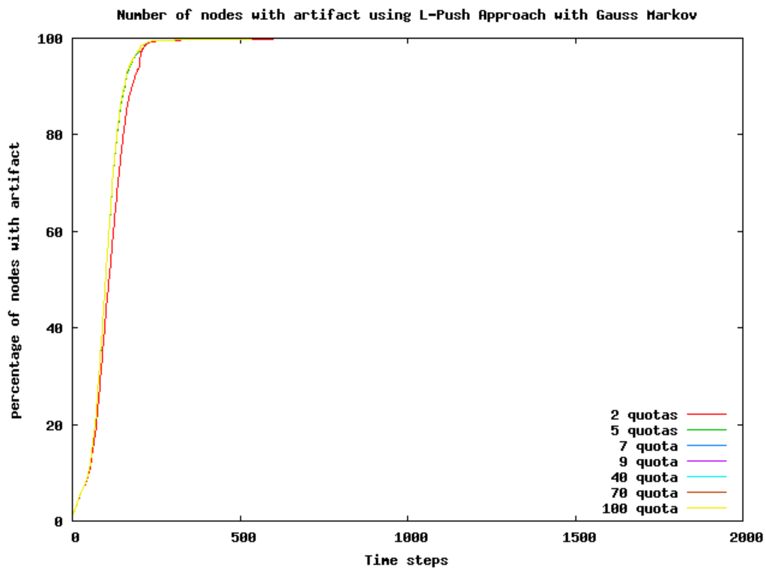


Figure 4.12. Average number of nodes with an artifact using L-Push Approach with Gauss Markov varying push quota

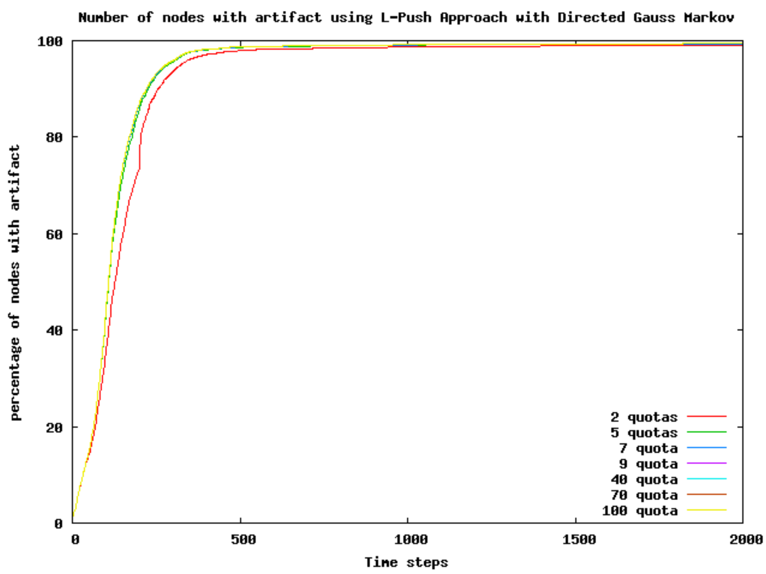


Figure 4.13. Average number of nodes with an artifact using L-Push Approach with Directed Gauss Markov varying push quota

has a higher chance of discovering information from other nodes. The nodes without information will keep querying until they discover the information. Therefore, the Push quota is only being used when a node that received a query has information. A small Push quota is sufficient to disseminate information effectively. This is shown in Table 4.6 where the average of final coverage values for all mobility models are the same even though the Push quota is increased.

- From Table 4.6, the highest overhead costs for the L-Push approach is 474.4 units which comes from D-GM model when Push quota is equal to 2. The lower overhead cost is 193.08 units which also comes from D-GM model when the Push quota is equal to 40. When the Push quota is set to a smaller number, for example Push quota is equal to 2, the number of Query generated is increased. This is because there are less nodes that discover information through the first query. So they keep generating queries until they find the information. Another interesting issue from this experiment is with the small number of Push quota, Information is disseminated equally as Push quota that is set to 100. This is due to the fact that within the interval time (20 seconds), the interactions between nodes are still going on. The interaction here means nodes keep sending queries to discover information. Therefore, within the interval time (20 seconds), the nodes still interact with each other until the next interval time, unless all nodes that are in range already have the information. This process make the L-Push different to the Greedy approach. In the Greedy approach, within the interval time the nodes stop interacting with each other when their Query quota is finished. That is why in Figure 4.2, 4.3, 4.4, and 4.5 we can see that the graph is not smooth as the query quota is increased.

4.4.3.2 The effect of Interval Time on Information spreading

Interval time is the duration of a node to renew its Push quota. For example if the Interval time is 20 seconds, then the push quota of a node will be reset to the original size of Push quota at every 20 seconds. This parameter is important to investigate because it determines the capability of a node to pass information to other nodes. We used 100 nodes with 50 time trials using four different mobility models for this experiment. In addition, we varied the Interval time so that we can determine the best performance Interval time to spread information. Table 4.7 shows result of the experiments.

Figures 4.14, 4.15, 4.16 and 4.17 show the effects of different Interval time on information spreading. From the figures, we can observe that:

- Different mobility models have different effect on information spreading. This is because the mobility models determine the possibility of node interactions which is a

Table 4.7. Statistics for different interval times with different Mobility Models using L-Push Approach.

Mobility Model	Interval Time	Average of final coverage	Standard Error	Standard Deviation	Accumulative Cost of overheads	
					Push	Query
Random Walk	5	88.489	0.268	25.449	97.6	97.6
	10	88.489	0.2682	25.449	97.6	97.6
	20	88.489	0.2683	25.449	97.6	97.6
	30	88.489	0.268	25.449	97.6	97.6
	60	88.489	0.2683	25.449	97.6	97.6
	90	88.487	0.268	25.450	97.6	101.28
	120	88.480	0.268	25.461	97.6	127.82
Random Waypoint	5	81.495	0.307	29.158	97.6	97.6
	10	81.495	0.307	29.158	97.6	97.6
	20	81.495	0.307	29.158	97.6	97.6
	30	81.495	0.307	29.158	97.6	97.6
	60	81.495	0.307	29.158	97.6	97.6
	90	81.493	0.307	29.16	97.6	100.38
	120	81.492	0.307	29.162	97.6	109.14
Gauss Markov	5	98.870	0.096	9.145	97.2	97.2
	10	98.870	0.096	9.147	97.2	97.2
	20	98.870	0.0964	9.147	97.2	104.0
	30	98.866	0.096	9.151	97.2	106.0
	60	98.866	0.096	9.151	97.2	106.04
	90	98.866	0.096	9.151	97.2	106.04
	120	98.866	0.096	9.151	97.2	106.04
Directed Gauss Markov	5	98.329	0.098	9.320	96.54	96.64
	10	98.328	0.098	9.325	96.54	97.88
	20	98.285	0.099	9.3677	96.54	119.76
	30	98.243	0.099	9.367	96.54	133.96
	60	98.241	0.099	9.368	96.54	137.9
	90	98.241	0.099	9.3675	96.54	137.9
	120	98.241	0.0987	9.367	96.54	138.22

compulsory requirement to have information dissemination. As we can see from Figures 4.14, 4.15,4.16 and 4.17, they show different rates of information dissemination. Guass Markov outperforms other mobility models and followed in order by D-GM, Random Walk and Random Waypoint.

- When increasing the interval time, there is no big change in the average number of nodes that receive information. This is due to the fact that the nodes are still interacting with each other until they discover the information. This happens because in

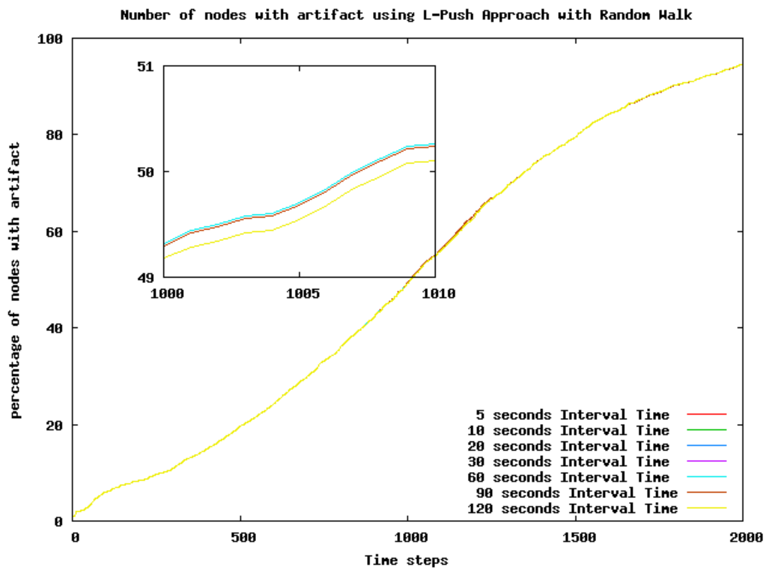


Figure 4.14. Average number of nodes having an artifact with different interval time using L-Push Approach and Random Walk while varying time interval

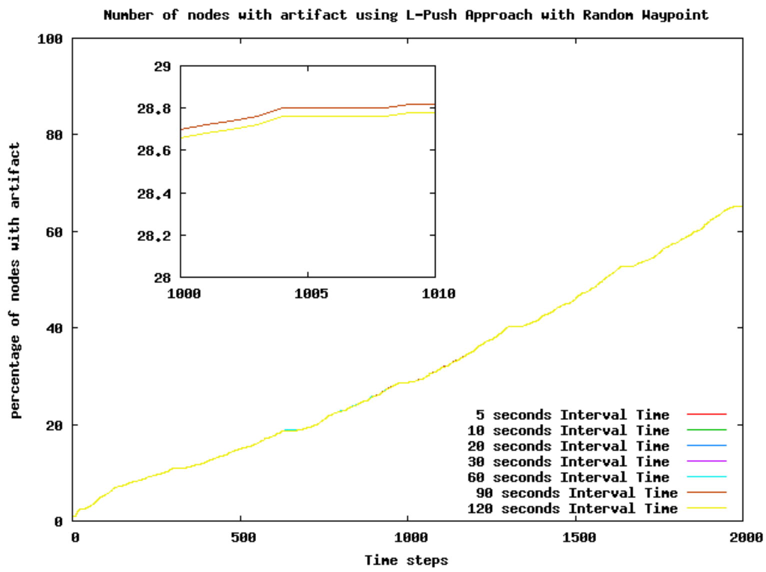


Figure 4.15. Average number of nodes having an artifact with different interval time using L-Push Approach with Random Waypoint while varying time interval

the L-Push a node has an unlimited Query quota in generating queries. Therefore, making the Interval time longer will not affect the number of nodes that have a Query quota as mention before that the nodes that looking for information are sending as much as they can do until they finds the information. What we learn from this approach is that the interval time is not affecting the information spreading.

- The overhead cost, Table 4.7 shows that the number of cost overheads is increasing

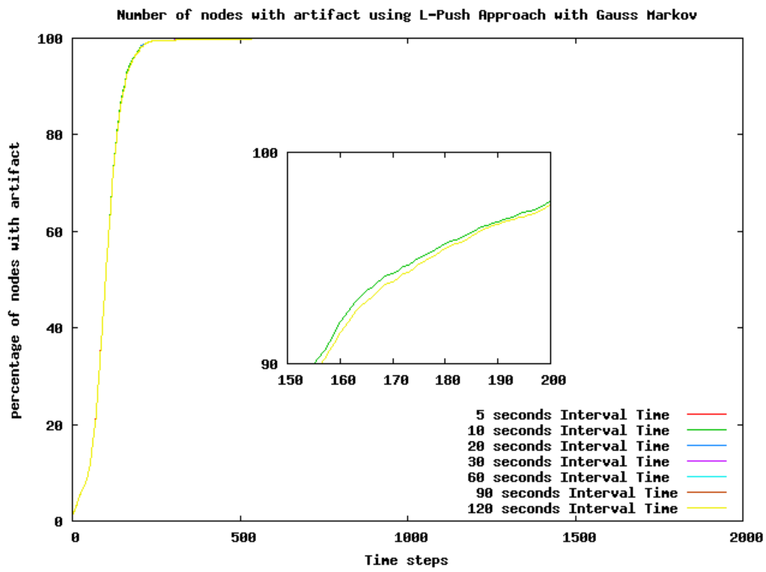


Figure 4.16. Average number of nodes having an artifact with different interval time using L-Push Approach with Gauss Markov while varying time interval

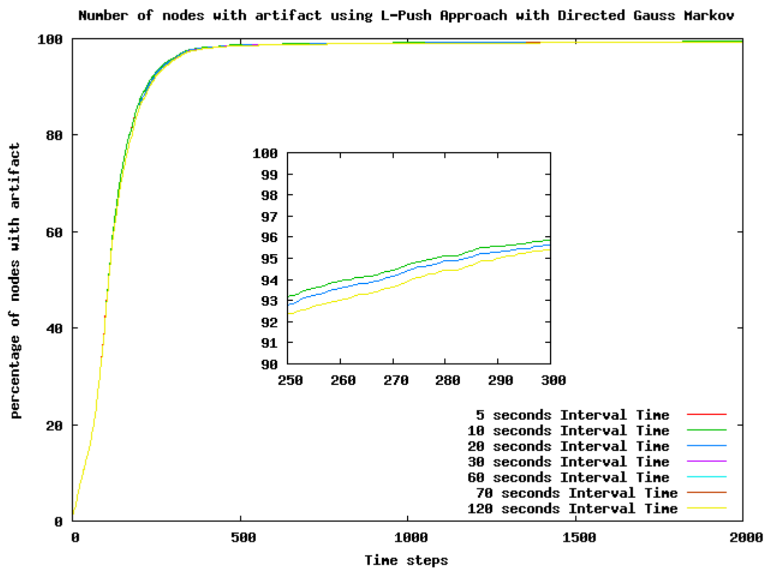


Figure 4.17. Average number of nodes having an artifact with different interval time using L-Push Approach with Directed Gauss Markov while varying time interval

when the interval time is increased. For example when using D-GM at 5 seconds Intervals, the overhead costs is 193.08 units. When the Interval time is increased to 120 seconds the overhead costs increases up to 234.76 units. This is due to the fact that the longer the Interval time the higher the possibility of nodes(that are looking for information) sending multiple queries to discover information. That is why the different overhead costs in the column accumulative cost of overheads is actually varied depending on the query overhead.

Based on the results presented in Section 4.4.3, we observed that to achieve **a high performance, combinations of Push quota and Interval time for the L-Push approach is achieved when the Push quota is equal to 5 and the Interval time is equal to 5 seconds using D-GM mobility model**. This selection is based on the amount of final coverage divide by the accumulative cost of overheads from Table 4.6 and 4.7. A high value from the division shows a high performance of L-Push approach.

4.4.4 Results-Spray and Relay

Spray and Relay works like the L-push approach but instead of renewing a quota in every time interval, the Push quota is fixed throughout the simulation and given at the beginning of simulation. The Push quota is not renewable but can be passed to other nodes. The size of the Push quota is the only factor that can varied the Spay and Relay performance besides the mobility models. In order to find the best performance of Push quota within 500 meters square to the 100 mobile nodes, we varied the size of quota for each experiment. We chose the spray quota between 5 to 100 using different mobility models. The results of each experiment are taken after 50 time trials of execution. Table 4.8 shows the results of the experiments.

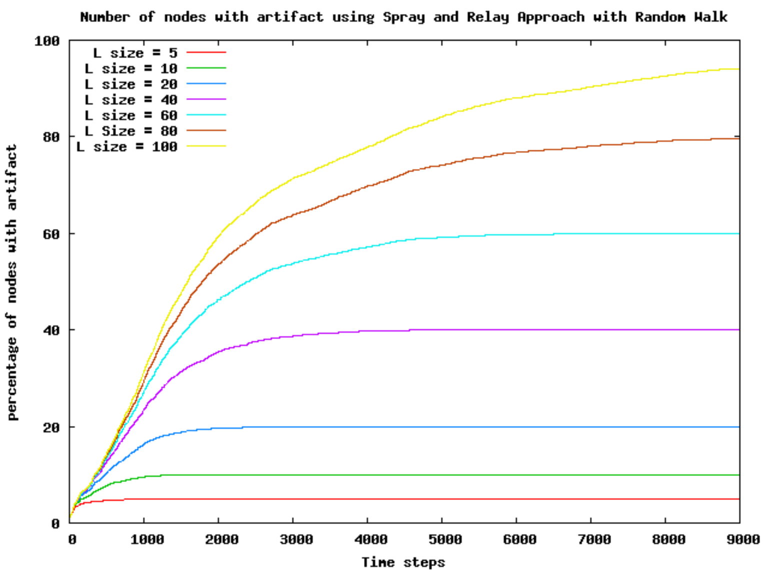


Figure 4.18. Average number of nodes have an artifact with different spray quota using Spray and Relay approach and Random Walk

Table 4.8. Statistics for different spray size with different Mobility Models using Spray and Relay approach.

Mobility Model	L size	Average of final coverage	Standard Error	Standard Deviation	Accumulative Cost of overheads	
					Push	Query
Random Walk	5	4.932	0.003	0.323	4.0	21547.22
	10	9.661	0.012	1.160	9.0	39398.42
	20	18.703	0.036	3.405	19.0	65613.26
	40	35.540	0.094	8.911	39.0	91434.36
	60	50.643	0.158	14.992	59.0	89578.64
	80	62.987	0.218	20.715	78.6	68332.7
	100	71.560	0.260	24.694	93.16	44846.32
Random Waypoint	5	4.906	0.004	0.358	9.0	36906.44
	10	9.478	0.014	1.341	9.0	36906.44
	20	17.860	0.043	4.041	19.0	57981.2
	40	31.808	0.109	10.334	38.94	71388.12
	60	42.371	0.173	16.415	57.84	60809.46
	80	49.209	0.214	20.333	71.38	50299.16
	100	53.453	0.239	22.667	71.38	42662.58
Gauss Markov	5	4.988	0.002	0.147	9.0	51705.54
	10	9.953	0.005	0.485	9.0	51705.54
	20	19.851	0.014	1.352	19.0	91278.06
	40	39.556	0.0359	3.401	39.0	136191.86
	60	59.131	0.061	5.754	59.0	135487.28
	80	78.474	0.088	8.392	79.0	91043.94
	100	97.016	0.120	11.430	98.98	6601.6
Directed Gauss Markov	5	4.992	0.001	0.129	9.0	21732.7
	10	9.967	0.004	0.410	9.0	21732.7
	20	19.874	0.013	1.194	19.0	37347.74
	40	39.565	0.034	3.203	39.0	53670.14
	60	58.967	0.059	5.584	58.98	51849.5
	80	77.481	0.087	8.256	78.44	32495.16
	100	91.936	0.114	10.842	95.18	7507.92

Figures 4.18, 4.19, 4.20 and 4.21 present the effect of spray quota on the information spreading when the spray is increased. From the figures we observe that:

- Different mobility models have different effect on the information dissemination. Gauss Markov and D-GM model have spread information rapidly at early stage before they achieve their optimum level respectively. This is can be observed from Figures 4.20 and 4.21. For Random walk and Random Waypoint as shown in Figure 4.18 and 4.19, the number of nodes with information are increased consistently as

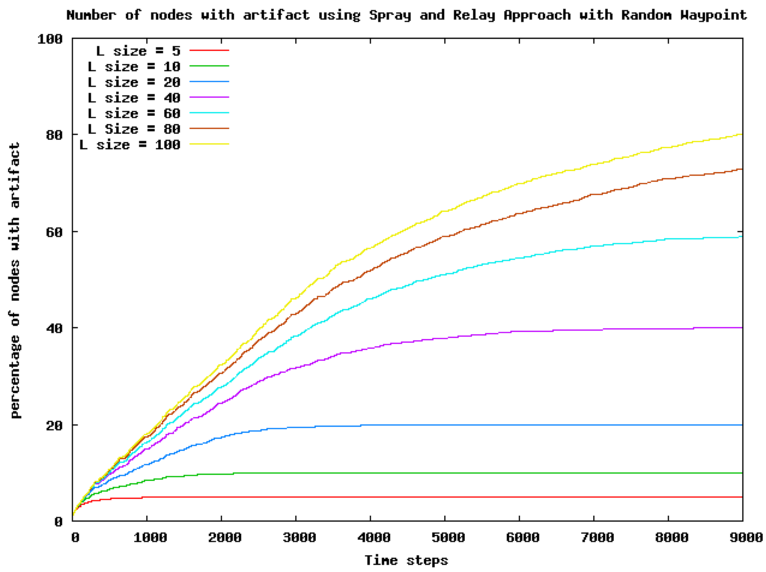


Figure 4.19. Average number of nodes have an artifact with different spray quota using Spray and Relay approach with Random Waypoint

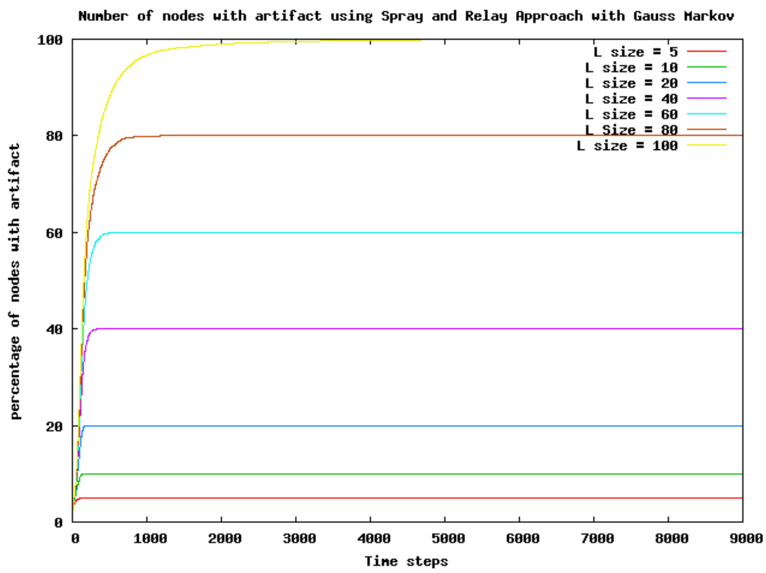


Figure 4.20. Average number of nodes have an artifact with different spray quota using Spray and Relay approach with Gauss Markov

the spray quota is increased. This is due to the fact that the more quota is assigned to a node, the more chances other nodes received information.

- Generally, when the spray (quota) size is increased the number of nodes that receive information are also increased. The size of spray is actually represents the capability of nodes to do push function. Therefore, a large size of spray quota increases the number of possible nodes that receive the information. Looking at Figures 4.20 and

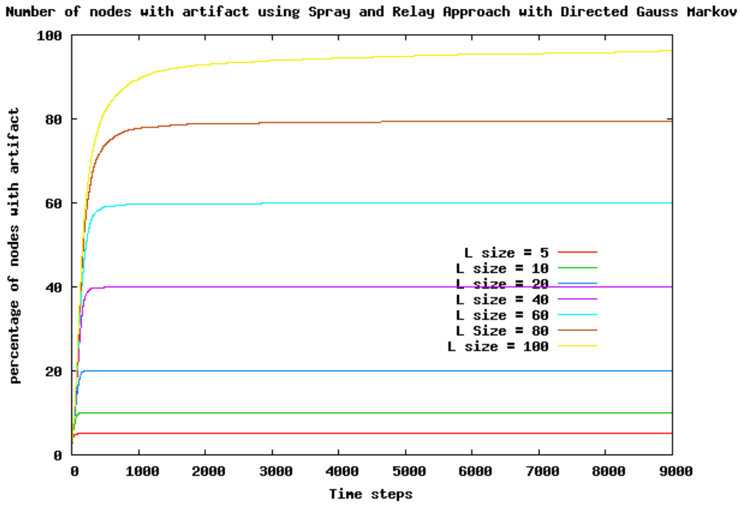


Figure 4.21. Average number of nodes have an artifact with different spray quota using Spray and Relay approach with Directed Gauss Markov

4.21, the size of spray represents the optimum number of nodes that actually will receive information.

From Table 4.8, Spray and Relay approach achieved its highest performance when the spray quota is 100 using the Gauss Markov mobility model. The average of final coverage nodes with artifact is approximately 97 percent with the accumulative cost of overheads at 6700.58 units.

4.5 Comparison of the best of each approach

In Table 4.9, we compiled the best combination of each approach from the experiment results best on the average of coverage and the accumulative overheads. We pick the lowest accumulative overheads with a reasonable performance for each dissemination approach respectively form Tables 4.2, 4.3, 4.6, 4.8.

Table 4.9. The best combination of each approach

Dissemination Approach	Mobility Model	Average of coverage at each time step	Standard Error	Standard Deviation	95 % C.I	Total accumulative Cost of Overheads (units)
Pure Push	D-GM	98.329	0.098	9.321	±0.192	212320.1
Greedy	Guass Markov	97.794	0.113	10.733	±0.222	1459.54
L-Push	D-GM	98.329	0.098	9.321	±0.193	193.18
Spray and Relay	Gauss Markov	97.016	0.120	11.430	±0.236	6700.58

From the Table 4.9, we observe that L-Push is the best approach because it has the lowest of total cost of overheads. More importantly it has the same average of final coverage as the Pure Push. Pure Push is the basic approach and it used as a benchmark approach or our dissemination approach.

Table 4.10 shows the best performance of each mobility model through out the experimentation results. We observe that the Gauss Markov and D-GM have very close performance in term of average of final coverage. In term of total accumulative overheads cost, D-GM has less cost compared to Gauss Markov. These facts indicate that both mobility models have very high nodes interactions.

Table 4.10. The best performance of each mobility model

Mobility Model	Push Approach	Push Quota	Average of final Coverage	Total accumulative Cost of Overheads (units)
Random Walk	L-Push	5	88.489	195.2
Random Waypoint	L-Push	5	81.494	195.2
Gauss Markov	L-Push	9	98.870	195.4
D-GM	L-Push	40	98.329	193.1

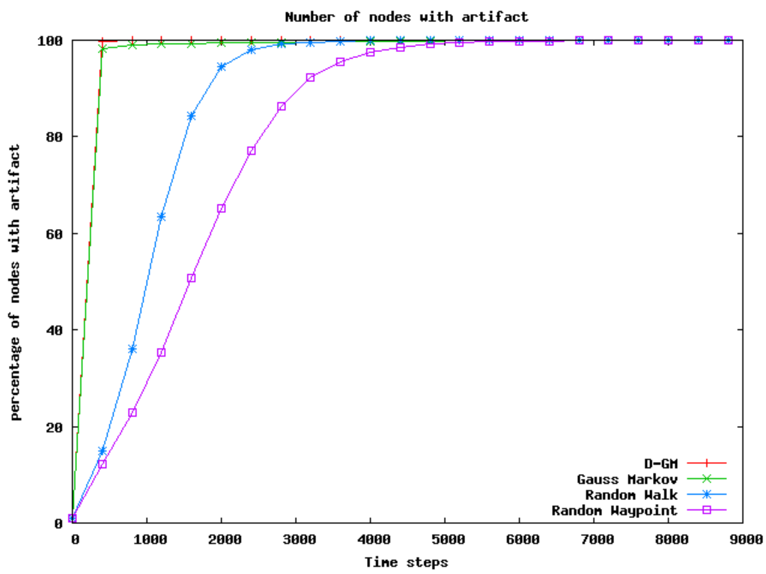


Figure 4.22. The best performance of each mobility model

From Figure 4.22, we can see that the D-GM and Gauss Markov disseminate information rapidly at the early stage (the line is overlapped each other). This is due to the fact that most of the nodes have high frequency interactions at early stage (0 - 50 seconds). For the Random Walk and Random Waypoint the data dissemination are increase steadily over the time as the movement of the nodes are determine randomly.

4.6 Conclusion

In this chapter we have presented four baseline data dissemination approaches (Pure Push, Greedy, L-Push, Spray and Relay). Experiments have been conducted with different sets of parameters to measure the performance of each approach on the information spreading. The performance is evaluated based on the overhead cost and the number of nodes that have information after the simulation time is finished.

Through the number of experiments in this chapter, we found out the push, query, quota and mobility model are all affecting the information spreading. The following has been identified as key points:

- Querying can be used effectively to locate an information. Asking every peer that is in range is the most effective way to discover information quickly. Of course at the early stage, most of the query is wasted, but as soon as the message spreads in the networks, the query is become more effective way to discover message. This is why L-Push has outperforms other approaches.

- Push is a fundamental part of information sharing. It is the most effective way to disseminate information but it causes a high duplication when push information blindly. Therefore, it needs to be controlled. We have learnt that query can be used to control the push mechanism. More importantly, with a good Push and Query combinations it can produce a good spreading information performance.

- Quota is a basic way of controlling the message duplication in push and query mechanism. However, it is very difficult to determine the optimum size of quota in order to achieve a good performance. The higher the quota the more chances of message spread quickly over the networks. However it always come with a higher push and query overheads.

- Different mobility models have different affect on the information spreading. The mobility model that encourages nodes to move actively to different locations diffuse message quickly. This is why Gauss Markov and D-GM models outperforms Random Walk and Random Waypoint mobility models in information spreading.

We have measured and presented the results in detailed of each approaches. We have shown the effect of query quota, push quota and interval time on the information spreading individually. Even though our approaches are simple(i.e. based on push, query and quota), we have shown that there is a potential to improve flooding effectiveness on information spreading. The following brief recap about dissemination approaches that used in this chapter.

- Pure push approach is the most basic dissemination approach. It only has a push functionality. A node that has information will always push information when in range with other nodes regardless whether the nodes have or does not have information. Because of that, Pure push is very efficient in information spreading but it has very high information duplication.

- Greedy approach is derived from Pure Push approach. Instead of just push the information to its peer as in Pure Push, Greedy approach push information (provided that node has information) if and only if there is query received. Therefore, we can adjust the information discover information rate by controlling the query frequency (query quota) for each node. The information spreading frequency is much dependent on the way the query is controlled.

- L-Push is the opposite of Greedy approach. Instead of limit the querying, it leaves the query open. It means that nodes can query every at any time. Further more, it also has limited push which control the information to be passed to it preferred peer. However, because of the information is homogenous and all nodes have the same interest in this chapter, the functionality of push is limited to the push information when is there is query. Consequently, adjusting the push size will not have the same effect as the node are freely to query to discover information.

- Spray and Relay is an idea that inspired from Spray and Wait protocol. The information in injected once in the network and leave the information spread through the nodes interactions. The unique of this approach is not only the information is pass but the number of quota is also being passed together with the information. The quota determines how many that the receiver is allowed to forward the information. This is actually indicates the potential number of node that will receive the information.

- The best approach in term of information dissemination is L-Push technique. The performance of L-Push is varies by using different mobility models. This is because the mobility model influence the frequency of nodes interactions. So, a combination of mobility model and push techniques is important key to achieve a high information dissemination performance.

The work in this chapter has built on the work of Chapter 3 which focuses on interaction between nodes in Mobile Peer to Peer networks. The work in Chapter 3 and this chapter tackle the foundation of data dissemination behavior in MP2P networks. However, it does not address any intelligent elements relating to the choice of who to prefer in terms of pushing information to or querying. This is a structural issue that we now consider. The next chapter focusing on proposing a social structure formation to guide the information dissemination over the P2P networks based on social relationship.

From this chapter also we can conclude that Push itself can achieve a high coverage performance with minimum query. The is based on the L-Push performance where even though it has a small number of Push quota, it has a high performance coverage. We will investigate further this concept in chapter 6 together with the social structure formation that is introduced in chapter 5.

Chapter 5

DIFFERENT TYPES OF SOCIAL STRUCTURE FORMATION

5.1 Introduction

In this chapter we look at the underlying structure that may emerge from repeated interaction between nodes caused by mobility models. By social structure formation we mean that nodes will come in range of each other on number of occasions. Depending on the mobility model certain nodes may more frequently see each other, and this can be modeled by a link in a graph, which when repeated across all nodes provides a *social structure*. A social structure has potential to be used for efficient information dissemination because it shows where data may be more likely to flow with reliability.

In the general literature, a social structure can represent many relationships between nodes (people, devices, organization) in social networks. The relationships between nodes can be made from different interdependencies, such as friendship, knowledge, beliefs and other elements that make nodes share or exchange things. According to Jhon Guare [12], people are actually separated by “six degrees of freedom”, which means that on average short paths connect any pair of people. This was originated by Millgram’s [31] experiment where people in United States are separated by about six people on average, which is referred to as a “small world”. Both studies show there are links connecting people in this world which they may not be aware of. It is based on this work, that we are inspired to examine if there are any social structures that can be formed using different mobility models based on how often nodes are in vicinity of each other.

In the previous works [6], [26], [52], mobility is used to form a social structure for data transfer. Here social links are represented by "utility functions" that are associated to the data objects and how frequently they group together. *To the best of our knowledge, there is no work that has made a comparison on the usefulness of different mobility models for different approaches of social structure formation (links between nodes).* This is an important contribution because examining the different ways in constructing social structures provides a clear picture of how social structure can be used to provide paths for data dissemination in opportunistic networks. With regards to social structure patterns, multiple paths in the graphs offer different routes for data to flow. However this introduces overhead. The other extreme is a tree structure, with minimal connectivity. We form a social structure based on a node's frequency of interactions for different mobility models. The social structure is constructed based on a node's current interactions which is updated locally by an individual node in different possible ways.

The purpose of investigating the social structure formation is to examine: (i) the difference between mobility models in terms of social structure produced; (ii) the effect of methods of forming links on the social structure observed. The social structure formation that we explore is based on the past interaction history as has been used in [11], [16], [27] [18] to form a social network. We introduce three different methods of constructing the social structure based on the nodes' past interactions. The methods are *Social structure based on average frequency interactions* (in Section 5.2.1), *Social Structure based on Periodicity Frequency Interactions* (in Section 5.2.2) and *Social Structure based on Sliding Window* (in Section 5.2.3).

5.2 Social Structure Formation Approaches

In this section we introduce three different approaches that we use to investigate the formation of social structure across different mobility models. The following are the terms that will be used in our approach:

- Period - assume n is a length of a period and t is the beginning of a new period. Therefore a period is $t + n$.
- Slot - is a unit of time step in a period.

The following sub sections provide a detail of each approach that is used to form a social structure in this chapter.

5.2.1 Social structure based on Average Frequency Interactions (AFI)

This approach uses the frequency of interactions to form a social structure. The "frequency interaction" measures how many times the same pair of nodes are co-located and interact throughout the simulation time. Each node records the interaction frequency with other nodes throughout the simulation. The average interactions frequency between node i and node j is determined by summing the number of times node i interacts with node j in all trials. The total value is divided by the number of trials to get the average interaction frequency between node i and j. A threshold is used to examine the frequency of interactions occurring between nodes. A ***Threshold*** *is a minimum value of frequency interaction for which a link is defined between two nodes.* For example if a threshold is x value, to establish a link between nodes, the nodes must interact more than or equal to x value over the duration of the simulation.

5.2.2 Social Structure based on Periodicity Frequency Interactions (PFI)

The formation of social structure using this approach is based on the interactions frequency that occur in a given period of time. For example, let the size of period be 5 minutes and x and y are mobile nodes. At the first period (0-5 minutes), assume that nodes x and y have managed to establish links between them. But, in the next period of time (5-10 minutes), no interaction is found between both nodes. Therefore, a link cannot be established between the nodes in the second period of time. This example shows that the nodes frequency interactions for different period affect the formation of social structure. Therefore, it is significant to investigate the impact of different period sizes on the social structure formation. Equation 5.2.1 is used to determine the percentage of a node that interact in a given period of time.

$$poe_{ij} = \frac{f_i}{nos} * 100\% \tag{5.2.1}$$

where poe_{ij} is percentage of times node i interacts with node j in a given period, where f_i is the total connections that exist between node i and j in the period and nos is the number slots in the period. The percentage of existence is used to determine the formation of links (social structure). In the experiment results, a threshold (in percentage) is used as a minimum percentage of existence for a particular node in a given period of time in order to form a social structures (links). For example a link between nodes is formed when a node's poe is higher that the threshold value. Based on Figure 5.1, let say node A has detected node B 300 times in period 1. So, the poe of node B is 10%. If the threshold is set to 10%, then a link from node A to B is establish at period 1. This link is only valid at period 1 and it will change depending on nodes A and B frequency interactions in the next period.

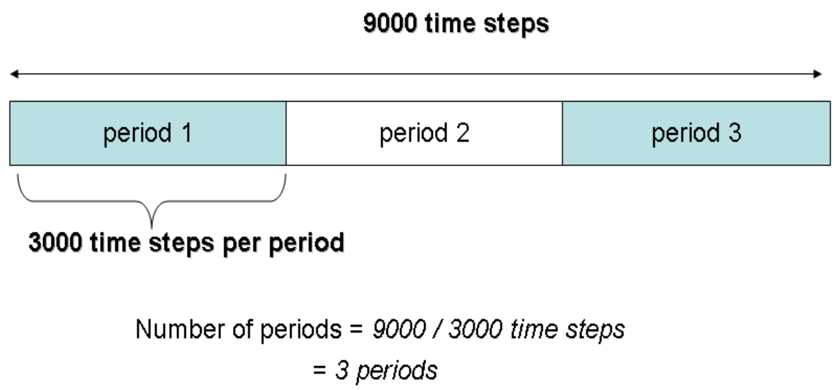


Figure 5.1. Number of periods for 3000 slots for 9000 simulation time.

5.2.3 Social Structure based on a Sliding Window Frequency Interaction (SWFI)

This approach uses a Sliding Window to determine a node interaction frequencies to form a social structure. Sliding Window (SW) is a frame that subdivided into number of slots. Each slot holds an ID of other nodes that interact with the node. The ID in each slot is shifted by one slot per time step. So, the contents of a frame changed over time. In our experiments, each node maintains its own SW locally. Each node records an ID of each node that it has established an interaction with. By compiling all information in SW, we can have a frequency of interaction for a particular node at the specific time.

The following definition are the terms that are used:

- Social Structure List (SSL) - records the nodes that are in a particular node's social structure.
- Social Structure Quota (SSQ) - is a maximum number of nodes that can be listed in SSL.
- Threshold - is a minimum frequency of a node found in a SW in order to be included in SSL.
- Link - is an edge between two nodes that are co-located more the threshold value.

Through this approach, we are able to capture nodes interaction frequency at a specific time. Figure 5.2 illustrates the SW mechanism. Based on the figure, the content of the slot is changed when a new input (node 6) is added into a frame. All contents are shifted by one at every time step. The last element in a frame will be dropped.

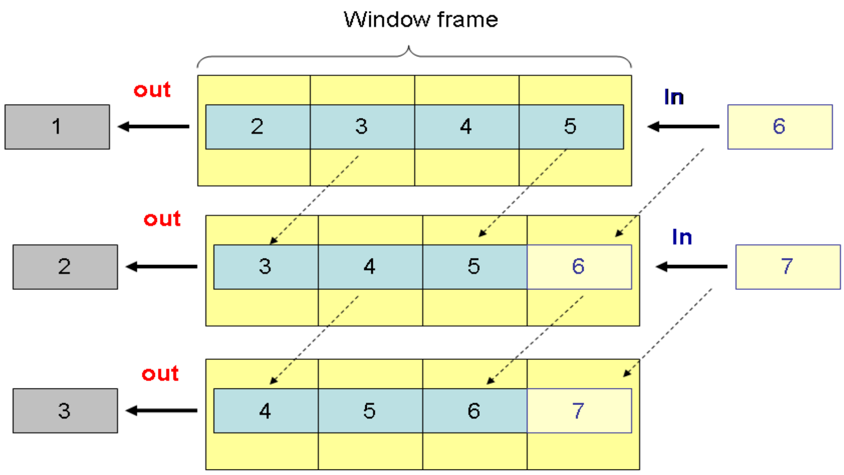


Figure 5.2. Sliding Window

The SW is updated at every time step. Assume that the current time step is 28 and the size of SW is 5. The observed steps within the SW are 27, 26, 25, 24 and 23. The nodes that listed in a SW are potentially to be included in a node *Social Structure List (SSL)*.

A link is formed when a particular node is found in the SW greater or equal to the threshold value. For example, let the size of SW be 40 and suppose node A has discovered node B 20 times in its SW. If the threshold is set equal to 10, then a social link is established between node A and B. However, the link between node A and B will be changed depending of the content of the Sliding Window after it has been shifted.

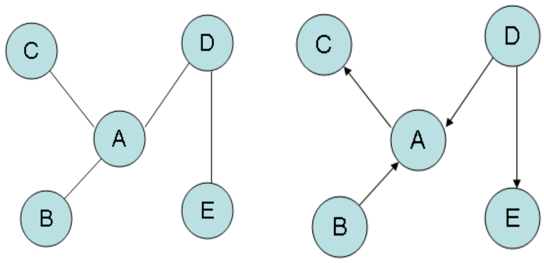
5.3 Experimentation

The organization of the experiments in this section is based on the social structure formation approaches as mentioned in Sections 5.2.1, 5.2.2, and 5.2.3. The main purpose of the experiments is to examine: (i) the different between mobility models in term of social structure produced (ii) the effect of methods of forming links on the social structure observed. The following sections are laid out with different parameters (according to the approach attributes) to examine different social structure pattern. The duration of simulation for every experiment is 9000 simulation time steps (15 minutes) and each simulation is repeated 50 times (i.e. trials) with different random seed generation (to avoid bias in the results). The test scenario considers 100 nodes randomly placed in a 500 meters x 500 meters region.

5.4 Key Performance Indicators (KPI's)

In social network analysis, there many metrics used to analyzed and understanding the roles of actors in social networks. In paper [3], different centrality measurements are presented that can be used to analyze social networks. In this chapter, we use the closeness centrality metric to evaluate our social structure. The closeness centrality measures how close a particular node is to all other nodes using the average shortest distance. According to [21], since closeness considers all pairs of nodes, it reflects the global connectivity of the social network structure.

Closeness centrality is designed to work on symmetric data, where each edge has no direction, as shown in Figure 5.3(a). Closeness centrality for symmetric data uses a single measurement to represent in and out degree of a node connections. However, in non-symmetric data, an edge between nodes has a direction as shown in Figure 5.3(b). A single measurement cannot represent a node connection to other nodes. This is because each link has a direction which shows the direction of the relationship between nodes. For example, node B can reach node C via node A but node C cannot reach node B because



(a) Symmetric (b) Non-Symmetric

Figure 5.3. Symmetric and non-symmetric relationship between nodes

there is no edge from node C to other nodes. Therefore, to measure the closeness for non-symmetric data, *Out Closeness* and *In Closeness* is used.

$$outCC_i = \left(\frac{N-1}{\sum_{j=1}^{N} d_{ij}} \right) \tag{5.4.1}$$

"Out closeness ($outCC_i$)" measures how close (in terms of distance) node i to other nodes. The $outCC_i$ can be calculated using Equation 5.4.1, where $i \neq j$ and d_{ij} is the length of the shortest path from node i to reach node j in a given network. In the case there is no path between node i and j, the distance between node i and j is set to the maximum distance i.e. the number of nodes that are found in the graph. A larger $outCC_i$ value indicates that nodes are very close to each other as the value of d_{ij} is smaller. So, a node that has a large $outCC_i$ value has a greater potential to disseminate data quickly.

$$inCC_i = \left(\frac{N-1}{\sum_{j=1}^{N} d_{ji}} \right) \tag{5.4.2}$$

"In closeness ($inCC_i$)" measures how close (in term of distance) from all nodes back to node i. Equation 5.4.1 is used to calculate the $inCC_i$ value for a particular node i, where $i \neq j$ and d_{ji} is the length of the shortest path from node j to reach node i in the networks. This metric measures how close other node to a particular node. A larger $inCC_i$ value indicates that a node has potential to receive information very quickly from other nodes.

Our data is non-symmetric since an edge has direction. The direction of an edge shows the flow of information. For example if node a forwards data to node b, then the edge direction is from node a to node b. If node b forwards data to node a, then the edge direction is from node b to node a. So, to measure both cases closeness centrality, we deploy $outCC_i$ and $inCC_i$ measurements in all our social structure results.

5.5 Result- Social structure based on Average Frequency Interactions (AFI)

This section presents the results of a social structure formation using the average frequency interactions (AFI) approach that has been explained in Section 5.2.1. Table 5.1 provides parameter settings that are used for the experiments. The following results discussion are organized based on the mobility models.

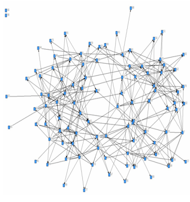
Table 5.1. Parameters Setting for API approach

Parameters	**Setting**
Number of nodes	100 nodes
Number of trials	50 times
Threshold	20,40,60,80,100,120,140
Mobility Model	Random Walk, Random Waypoint, Gauss Markov and D-GM

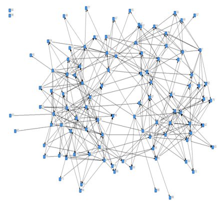
5.5.1 AFI social structure formation using Random Walk

Based on Figure 5.4, we can observe that varying the threshold value changes the social structure patterns. This is because the threshold affects the formation of social structure. It is also very rare to have nodes that can maintain a high number of interactions frequency between the same nodes, especially in the mobile environment. Therefore, as the threshold increases, the social structure density is become less dense. For example when threshold is set between 20 to 60 (Figure 5.4(a), 5.4(b) and 5.4(c)), no significant changes in the social structure density can be seen. This is because most of the nodes are seeing each other averagely 60 times. So, setting the threshold less than 60 will results in a smaller differences in social structures density. However, when the threshold is set to 80 and above, the density of the social structure is less as shown in Figure 5.4(d), 5.4(e) and 5.4(f).

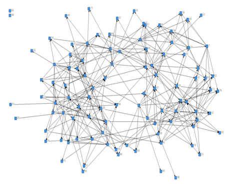
Table 5.2 shows the in and out closeness centrality. As we can see, a lower threshold value has an average a high closeness centrality value which means nodes are close to other in the network. This gives us a clue that it is useful to dissemination information with a



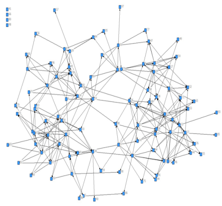
(a) Threshold >=20



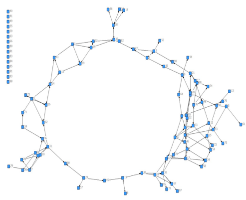
(b) Threshold >=40



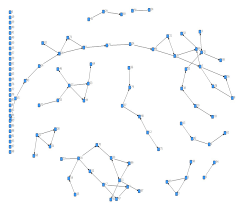
(c) Threshold >=60



(d) Threshold >=80



(e) Threshold >=100



(f) Threshold >=120

Figure 5.4. AFI Social Structure using Random Walk Model with Different Thresholds

Table 5.2. Closeness Centrality of AFI Social Structure using Random Walk

Threshold	**Average of Closeness centrality**	
	In Closeness	Out Closeness
20	1.594	8.129
40	1.594	8.129
60	1.594	8.129
80	3.100	4.014
100	1.043	1.046
120	1.008	1.008
140	1.008	1.008

lower threshold value. Note that, the unusual values of In Closeness and Out Closeness for threshold=80 cannot be explained.

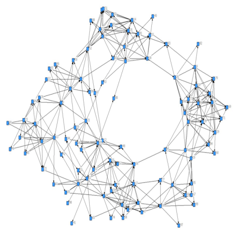
5.5.2 AFI Social Structure Formation using Random Waypoint

Using the Random Waypoint mobility model, we observe that nodes are co-located more frequently. Even the threshold value is bigger, the social structure density is better as compared to the Random Walk.

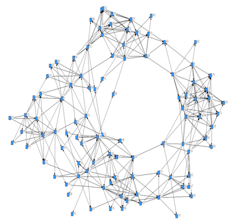
In Figure 5.5, the pattern of social structures in figures 5.5(a), 5.5(b) and 5.5(c) remain the same even though the threshold value is increased. This indicates that the nodes co-located between 20 to 80 times.

As the threshold is further increased, the number of links between nodes is decreased. This is can be seen from Figure 5.5(e) where only a few nodes are connected compared to 5.5(a). As the threshold is further increased (Figure 5.5(g)), the links between nodes are clearly disconnected. This is because it is hard for nodes to be co-located with the same node frequently at the same time.

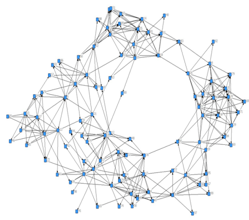
Looking at the statistics given in Table 5.3, the number of links (in degree and out degree) and the closeness centrality values are better when the threshold value is small. This is because it is easy to form a relationship among the nodes with a small number encounter frequency between nodes. This indirectly implies that less restriction in forming social structures improves the performance of dissemination in the network.



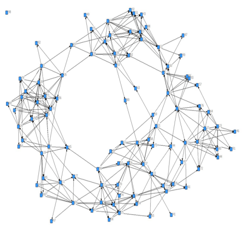
(a) Threshold >=20



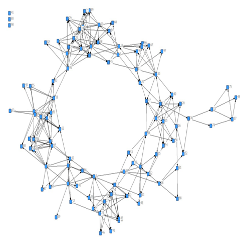
(b) Threshold >=40



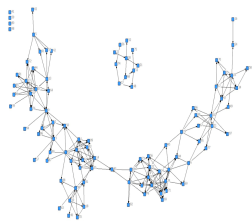
(c) Threshold >=60



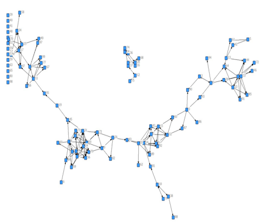
(d) Threshold >=80



(e) Threshold >=100



(f) Threshold >=120



(g) Threshold >=140

Figure 5.5. AFI Social Structure using Random Waypoint with Different Thresholds

Table 5.3. Closeness Centrality of AFI Social Structure using Random Waypoint

Threshold	Average of Closeness centrality	
	In Closeness	Out Closeness
20	1.375	4.176
40	1.375	4.176
60	1.375	4.176
80	1.373	3.655
100	1.318	1.642
120	1.070	1.080
140	1.041	1.045

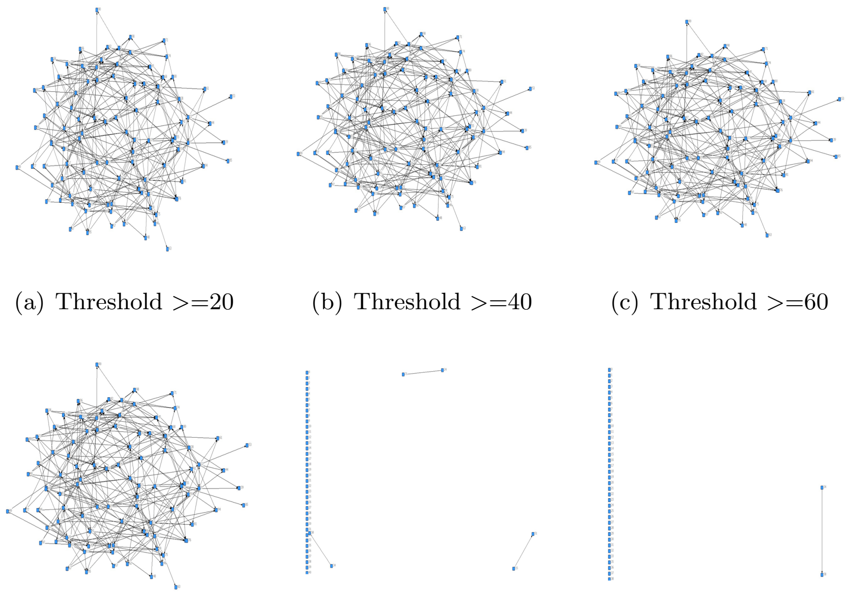
5.5.3 AFI Social Structure Formation using Gauss Markov

The Gauss Markov mobility model offers a high frequency of nodes interaction compared to Random Walk and Random Waypoint. Focusing on Figure 5.6, when the threshold is increased from 20 to 60, there are no significant changes found in the social structure as shown in Figures 5.6(a), 5.6(b) and 5.6(c). However, when the threshold is set to 100, only a few social structures are found as shown in Figure 5.6(e). This means that Gauss Markov mobility model potentially can give a better data dissemination performance when the threshold is less than 100.

From the statistics given in Table 5.4, we observe that with a lower threshold value, nodes averagely have a high closeness centrality value. This is shows that there are number of short paths in order to reach other nodes in the network. This scenario shows that we can disseminate data very quickly by using a small threshold value.

Table 5.4. Closeness Centrality of AFI Social Structure based on using Gauss Markov

Threshold	Average of Closeness centrality	
	In Closeness	Out Closeness
20	1.603	10.678
40	1.603	10.678
60	1.603	10.678
80	1.002	1.002
100	1.000	1.000
120	1.000	1.000
140	1.041	1.045



(a) Threshold >=20 (b) Threshold >=40 (c) Threshold >=60

(d) Threshold >=80 (e) Threshold >=100 (f) Threshold >=120

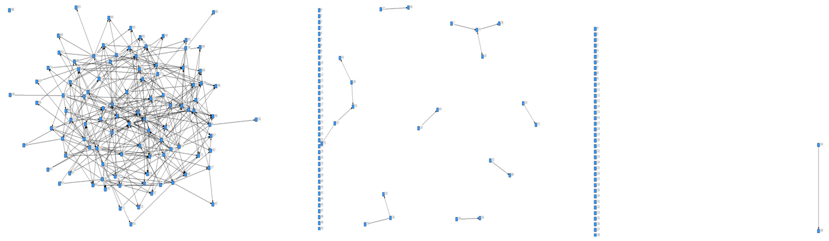
Figure 5.6. AFI Social Structure using Gauss Markov Model with Different Thresholds

5.5.4 AFI Social Structure Formation using D-GM

As expected, the social structure of a node using D-GM mobility model is better when a small threshold is used.

As we can see from Figure 5.7, increasing the threshold value from 20 to 40 changes the social structure density drastically. This is because not many nodes are able to maintain co-location for a long period of time with the same nodes. Moreover, the D-GM mobility model prevents nodes to move to the previous nodes' position. Therefore, the chance of meeting the same nodes is very low.

Looking at the statistics given in Table 5.5, the average value of closeness centrality is better when a threshold is set to a lower value. This means that there are number of shortest links between nodes which potential can be used to disseminate information quickly to all nodes in the network.



(a) Threshold >=20 (b) Threshold >=40 (c) Threshold >=60

Figure 5.7. AFI Social Structure using D-GM with Different Thresholds

Table 5.5. Closeness Centrality of AFI Social Structure based on using D-GM

Threshold	Average of Closeness centrality	
	In Closeness	Out Closeness
20	1.494	3.104
40	1.002	1.002
60	1.000	1.000

5.6 Result- Social Structure Based on Periodicity of Frequency Interactions (PFI)

In this section, we explore the construction of social structures using a the periodicity of frequency interactions approach which described in Section 5.2.2. We use three different size of periodicity i.e. 300 seconds, 600 seconds and 3000 seconds. Each size of periodicity is examined using different mobility models to analyze the impact on the social structure formation. Figure 5.6 provides the parameter settings that are used for the experiments. The threshold is used as lower bound of the percentage of nodes that interact in a give period. A link between nodes will be established when the interaction percentage is above the threshold.

Table 5.6. Parameters Setting for PFI approach

Parameters	Setting
Number of nodes	100 nodes
Threshold	10% and 25%
Periodicity	300s, 600s and 3000s
Mobility Model	Random Walk, Random Waypoint, Gauss Markov and D-GM

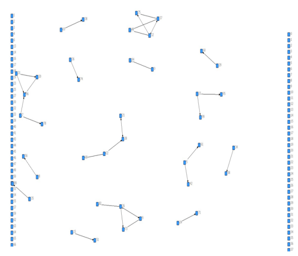
5.6.1 PFI social Structure Formation using Random Walk

From Figure 5.8, we cannot see any social structures are formed at 300 seconds. This is because 300 seconds is too short for a node to discover other nodes and establish relationships. Increasing the period size gives more time and opportunity for a node to be co-located and develop a social structure. This can be seen in Figures 5.10(a) where more links are found between nodes when the periodicity size is increased.

From the statistics in Table 5.7, we can see that the average value of closeness centrality is increased when the size of period is increased from 600 seconds to 3000 seconds. This is gives nodes more chance to be co-located and discover each other within the same period. This indirectly provides more ways to reach other nodes in the network.

(a) Threshold >=10 % (b) Threshold >=25 %

Figure 5.8. PFI Social Structure for Random Walk with periodicity = 300 seconds

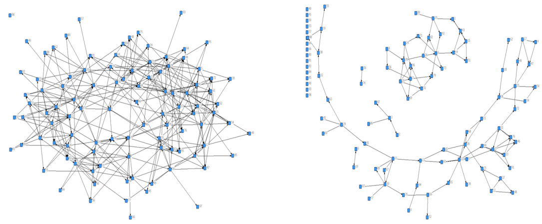


(a) Threshold >=10 % (b) Threshold >=25 %

Figure 5.9. PFI Social Structure for Random Walk with periodicity = 600 seconds

5.6.2 PFI Social Structure Formation using Random Waypoint

From Figure 5.11, a social structure is easily formed when the threshold is small. This is because the formation of relationship between nodes is less strict, i.e. the nodes only have



(a) Threshold >=10 % (b) Threshold >=25 %

Figure 5.10. PFI Social Structure for Random Walk with periodicity = 3000 seconds

Table 5.7. Closeness Centrality of PFI Social Structure using Random Walk

Period	**Threshold (%)**	**Average of Closeness centrality**	
		In Closeness	Out Closeness
300	10	0.000	0.000
	25	0.000	0.000
600	10	1.004	1.004
	25	1.011	1.476
3000	10	1.600	9.970
	25	1.028	1.029

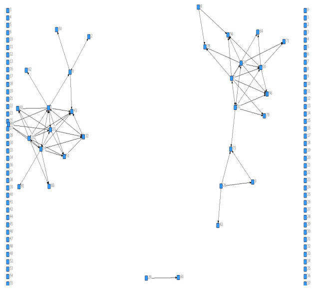
to be co-located for a short time only.

Comparing the number of links found in Figure 5.11 and Figure 5.12 the increased periodicity in Figure 5.12 gives more links and is more dense. This is because the nodes have a long period of time which gives a chance for nodes to establish relationships with other nodes. So, as we expected, with periodicity at 3000 more links can be formed as shown in Figure 5.13.

The statistics in Table 5.8 shows the average value of closeness centrality is improved when the periodicity size is increased. This is due to the fact that it is easy to maintain a relationship among the nodes in a long period of time which leads to the establishment of many links between different nodes. This makes the nodes closer to each other.

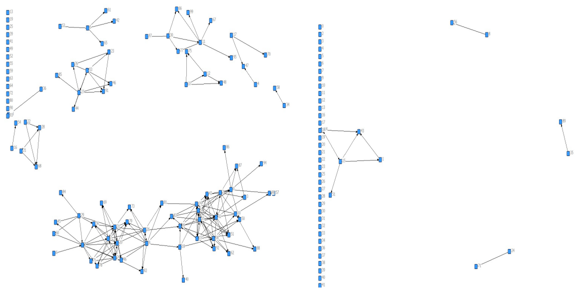
Table 5.8. Closeness Centrality of PFI Social Structure using Random Waypoint

Period	**Threshold (%)**	**Average of Closeness centrality**	
		In Closeness	Out Closeness
300	10	1.015	1.017
	25	0.000	0.000
600	10	1.058	1.072
	25	1.002	1.002
3000	10	1.220	2.973
	25	1.055	1.060



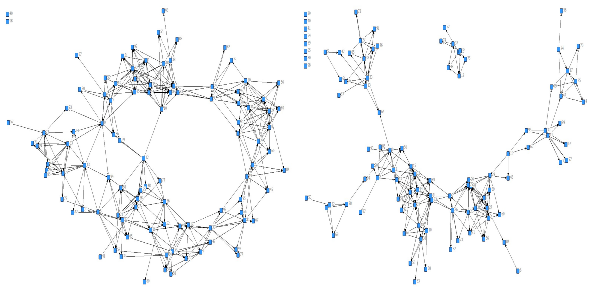
(a) Threshold >=10 % (b) Threshold >=25 %

Figure 5.11. PFI Social Structure for Random Waypoint with periodicity = 300 seconds



(a) Threshold >=10 % (b) Threshold >=25 %

Figure 5.12. PFI Social Structure for Random Waypoint with periodicity = 600 seconds



(a) Threshold >=10 % (b) Threshold >=25 %

Figure 5.13. PFI Social Structure for Random Waypoint with periodicity = 3000 seconds

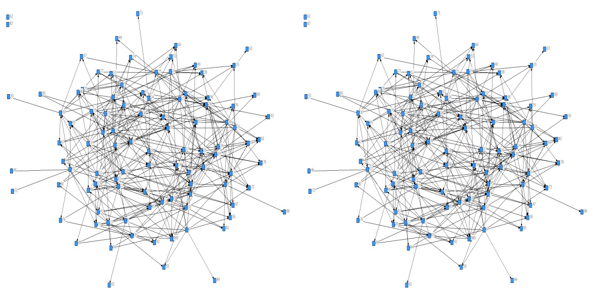
5.6.3 PFI Social Structure Formation using Gauss Markov

A social structure generated using the Gauss Markov mobility model has better characteristics than the Random Walk and Random Waypoint mobility models. Figure 5.14(a) shows that even though the period of time used very small, a social structure can still be traced. This is because the Guass Markov mobility model determines the next location of a node based on the current node's position not randomly as in Random Walk and

Random Waypoint mobility model. This appears to increase the chance of co-location.

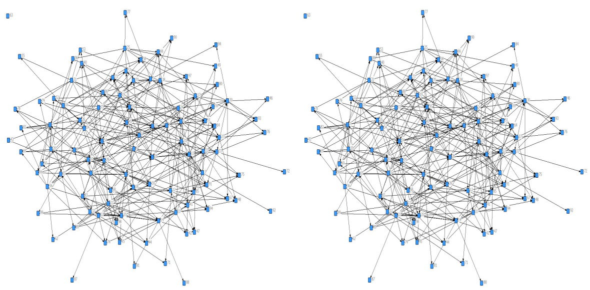
As in the Random Walk and Random Waypoint mobility model experiments, increasing the threshold value effects the formation of social structure. This is because the threshold value determines the social formation of links between nodes.

Looking at the statistics in Table 5.9 the centrality closeness nodes is identical for both thresholds. It is then faster to disseminate information to other nodes in the network because nodes have different links to reach to different nodes.



(a) Threshold >=10 % (b) Threshold >=25 %

Figure 5.14. PFI Social Structure for Gauss Markov with periodicity = 300 seconds

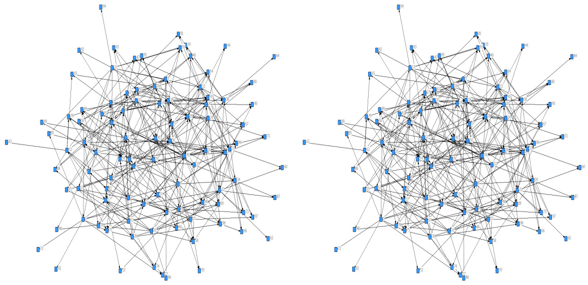


(a) Threshold >=10 % (b) Threshold >=25 %

Figure 5.15. PFI Social Structure for Gauss Markov with periodicity = 600 seconds

Table 5.9. Closeness Centrality of PFI Social Structure using Gauss Markov

Period	**Threshold (%)**	**Average of Closeness centrality**	
		In Closeness	Out Closeness
300	10	1.485	3.297
	25	1.485	3.297
600	10	1.535	3.910
	25	1.535	3.910
3000	10	1.610	14.236
	25	1.610	14.236



(a) Threshold >=10 % (b) Threshold >=25 %

Figure 5.16. PFI Social Structure for Gauss Markov with periodicity = 3000 seconds

5.6.4 PFI Social Structure Formation using D-GM

From Figure 5.17(a) we observe that the structure between nodes is very weak. It is notable that the D-GM mobility model reduces the possibility discovering each other.

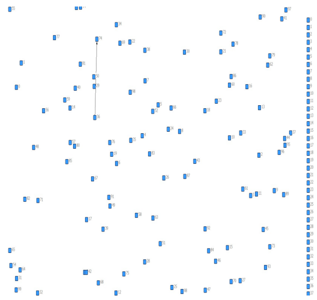
However, when the time period is increased to 600 seconds, the number of links between nodes is increased. Figure 5.18(a) shows a social structure formation when using a lower threshold value. When the threshold is further increased, nodes do not manage to maintain a link to other nodes because it is hard for a node to be co-located with the same nodes more frequently.

Generally, increasing the size of periodicity gives sufficient time for a node to establish social relationships with other nodes. This is shown in Figure 5.19 which has more links as compared to Figure 5.17 and 5.18

Based on the statistics given in Table 5.10, we can observe that the average value of closeness centrality increases when the period size is increased. This is because in a long period of time, nodes manage to established many relationship with other nodes and link density increases.

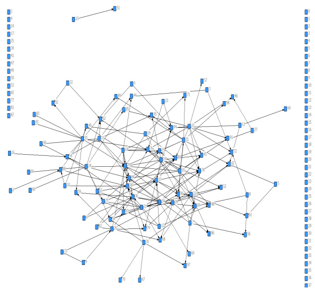
Table 5.10. Closeness Centrality of PFI Social Structure using D-GM

Period	**Threshold (%)**	**Average of Closeness centrality**	
		In Closeness	Out Closeness
300	10	1.000	1.000
	25	0.000	0.000
600	10	1.048	1.052
	25	0.000	0.000
3000	10	1.222	1.282
	25	1.222	1.282



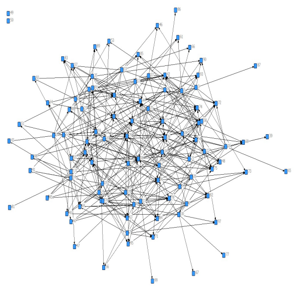
(a) Threshold >=10 % (b) Threshold >=25 %

Figure 5.17. PFI Social Structure for D-GM with periodicity = 300 seconds



(a) Threshold >=10 % (b) Threshold >=25 %

Figure 5.18. PFI Social Structure for D-GM with periodicity = 600 seconds



(a) Threshold >=10 %

Figure 5.19. PFI Social Structure for D-GM with periodicity = 3000 seconds

5.7 Result-Social Structure Based on Sliding Window Frequency Interactions (SWFI)

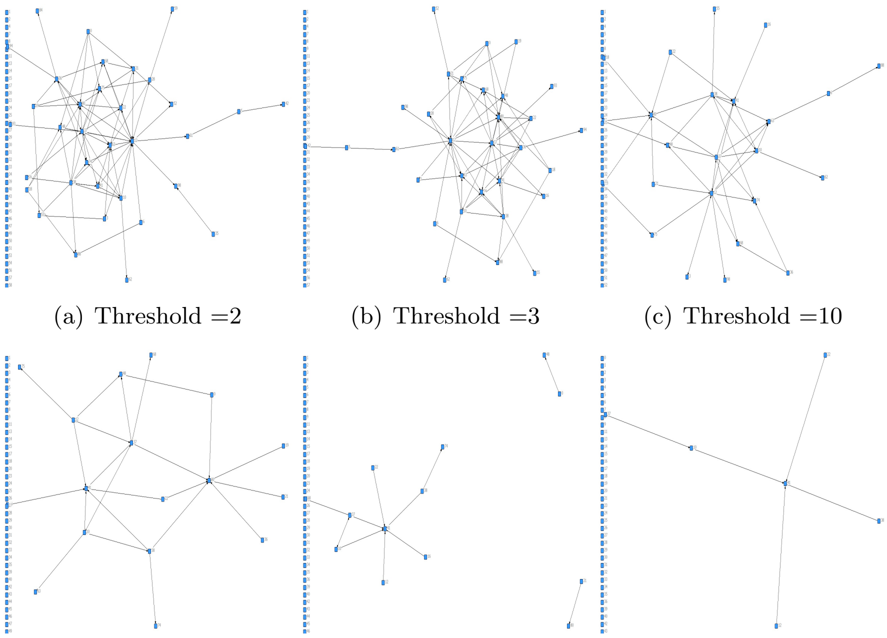
The results presented in this section are constructed based on Sliding Window Frequency Interaction (SWFI) approach as explained in Section 5.2.3. Table 5.11 provides parameters that use for the simulations.

Table 5.11. Parameters Setting for SWFI approach

Parameters	**Setting**
Number of nodes	100 nodes
Number of trials	50 times
Social structure quota	10 members
Sliding Window Size	40
Threshold	2,3,10,20,30,40

5.7.1 SWFI Social Structure using Random Walk

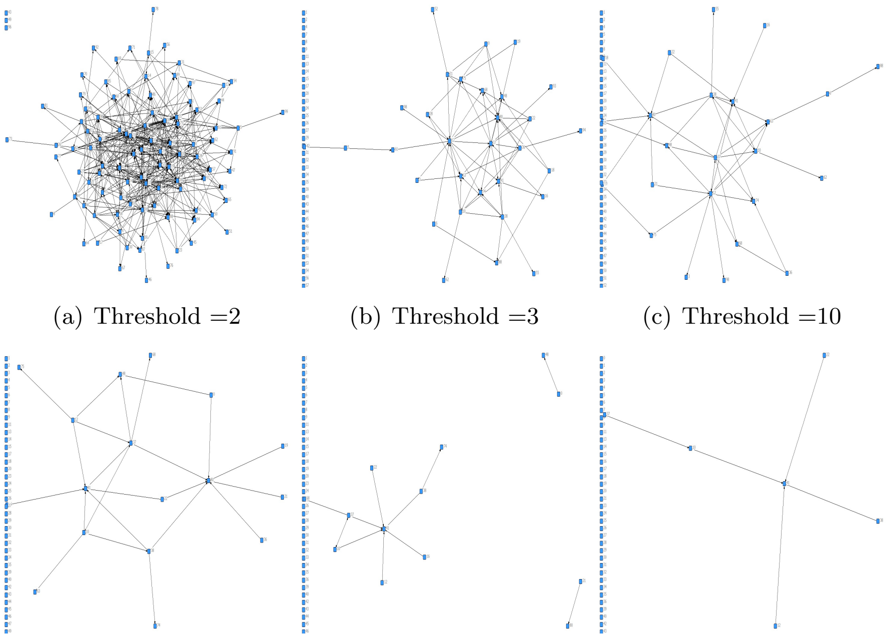
From Figure 5.20, we can observe that the number of links between nodes decreases when the threshold value is increased. This is because it is difficult for a node to be co-located with the same nodes within in a short period of time. As can be seen in Figure 5.20(f), when threshold increases, the social structure between nodes degrades. This is because it is hard for a node be co-located frequently with the same nodes in a short period of time.



(a) Threshold =2 (b) Threshold =3 (c) Threshold =10
(d) Threshold =20 (e) Threshold =30 (f) Threshold =40

Figure 5.20. SWFI Social Structure formation after 500 time steps simulation time using Random Walk

Increasing the period of time from 500 to 1000 time steps change the nodes' social structure. This is mainly because nodes have more opportunity or time to meet other nodes which indirectly enrich the density of social relations between nodes. This is can be observed when we compare the social structure in Figure 5.20(a) and Figure 5.21(f) which respectively using 500 and 1000 time steps.



(a) Threshold =2 (b) Threshold =3 (c) Threshold =10

(d) Threshold =20 (e) Threshold =30 (f) Threshold =40

Figure 5.21. SWFI Nodes Structure after 1000 time steps using Random Walk

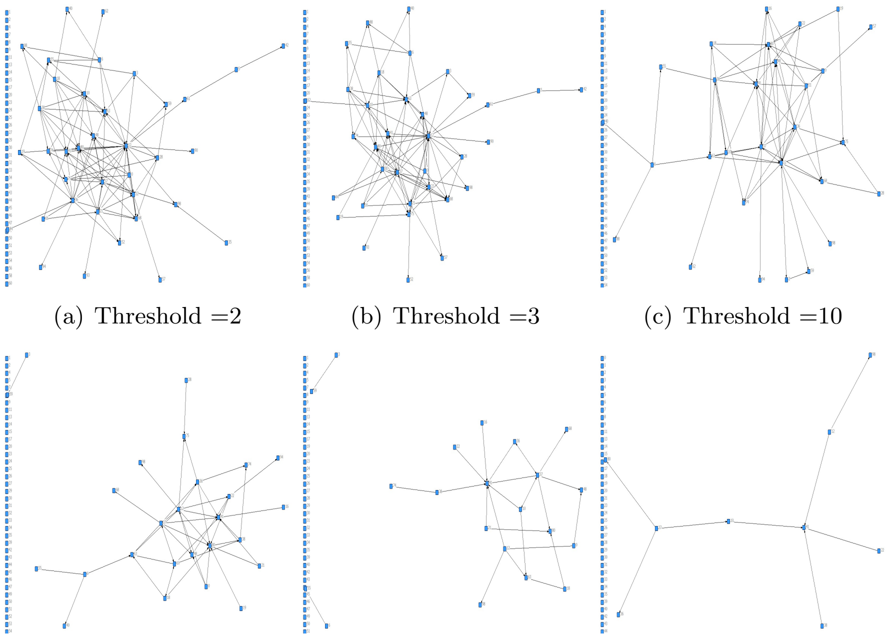
Looking at the statistics in Table 5.12 we found that the closeness centrality value increases when the social structure is captured after 1000 time steps. From a data dissemination perspective, this is shows that it is faster to diffuse information after 1000 steps rather than after 500 time steps.

Table 5.12. Closeness Centrality of SWFI Social Structure using Random Walk

time steps	**Threshold (%)**	**Average of Closeness centrality**	
		In Closeness	Out Closeness
500	2	1.041	1.048
	5	1.031	1.035
	10	1.011	1.012
	20	1.005	1.005
	30	1.001	1.001
	40	1.001	1.001
1000	2	1.408	2.960
	5	1.385	2.712
	10	1.306	1.818
	20	1.186	1.359
	30	1.056	1.074
	40	1.011	1.011

5.7.2 SWFI Social Structure using Random Waypoint

Random Waypoint mobility is variation of Random Walk mobility model. It moves randomly and stop at certain location for a period of time before move to the next destination. Thus, it limits the nodes interact with different nodes. Based on Figure 5.22, increasing the threshold value decreases the number of links. This is because it is difficult for nodes to maintain a relationship with the same nodes as both of them must be in range frequently.



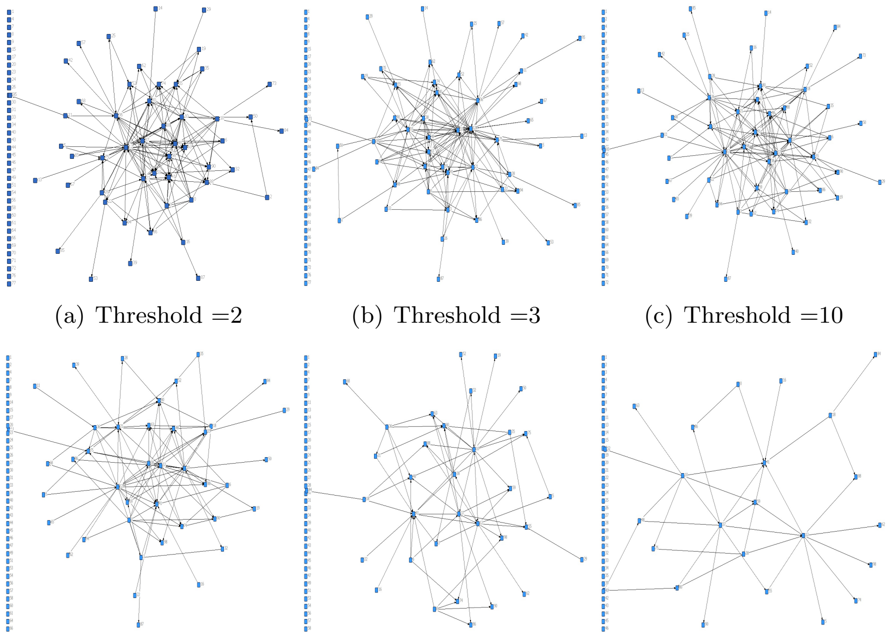
(a) Threshold =2 (b) Threshold =3 (c) Threshold =10

(d) Threshold =20 (e) Threshold =30 (f) Threshold =40

Figure 5.22. Nodes Social Structure after 500 time steps using Random Waypoint

As in the previous experiments, capturing a social structure after 1000 time steps simulation has a different pattern as compared to social structure pattern after 500 seconds. A greater number of links are found after 1000 time steps. This is due to the fact that nodes have more time to discover different nodes in a long period of time. Figure 5.23 shows the social structure that capture after 1000 time steps.

Looking at Table 5.13, the average closeness centrality value between nodes are not much different and the closeness centrality value is very small. This implies that disseminating information under the Random Waypoint mobility model is not particularly effective.



(a) Threshold =2 (b) Threshold =3 (c) Threshold =10

(d) Threshold =20 (e) Threshold =30 (f) Threshold =40

Figure 5.23. SWFI Social Structure after 1000 steps using Random Waypoint

Table 5.13. Closeness Centrality of SWFI Social Structure using Random Waypoint

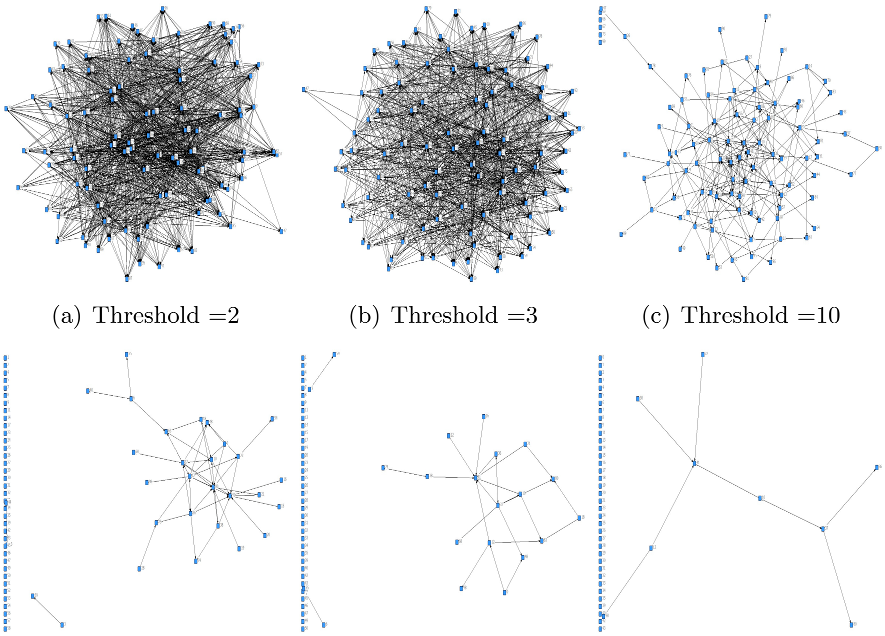
time steps	**Threshold (%)**	**Average of Closeness centrality**	
		In Closeness	Out Closeness
500	2	1.050	1.063
	5	1.043	1.053
	10	1.028	1.032
	20	1.013	1.014
	30	1.004	1.004
	40	1.001	1.001
1000	2	1.408	2.960
	5	1.098	1.145
	10	1.095	1.139
	20	1.051	1.064
	30	1.022	1.025
	40	1.007	1.008

5.7.3 Social Structure using Gauss Markov

The Gauss Markov mobility model determines the next node's position based on the current node's location. This mechanism helps nodes to discover the same nodes more frequently as compared to random assignment that implemented in Random Walk and Random Waypoint mobility model.

In Figure 5.24, we can see many links are formed between nodes when the threshold is small. This is because it is easy for nodes to be part of another nodes social network, since a node only needs to be co-located with same node twice or more within the same Sliding Window.

In Figure 5.25, we can see that all nodes are connected to the network when the threshold is small. This actually increases the accessibility to reach other nodes very quickly. As we can see in Table 5.14, the average value of closeness centrality is very high which shows that there is strong links between nodes. Therefore, disseminating information from one node to another can be done very quickly when a structure is defined with a small threshold.



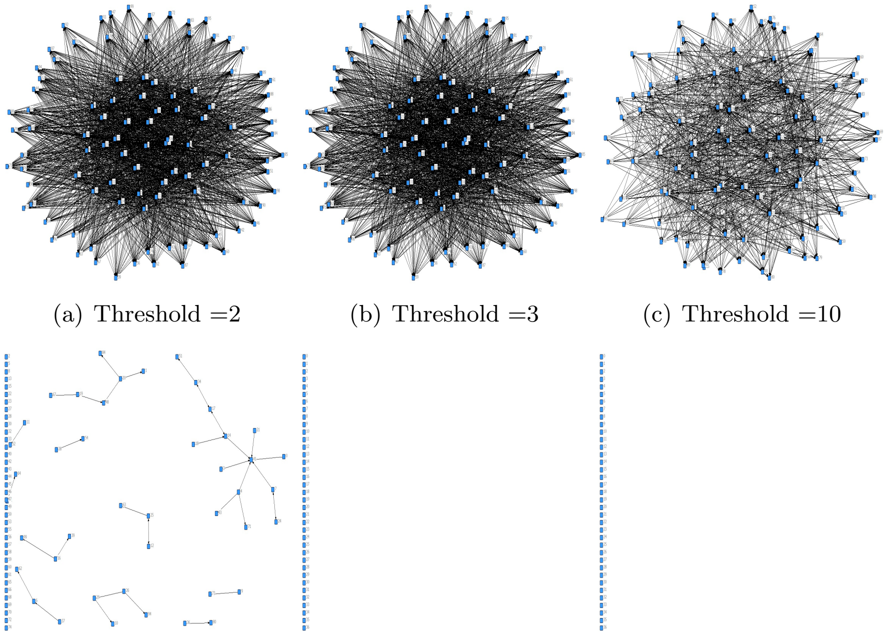
(a) Threshold =2 (b) Threshold =3 (c) Threshold =10

(d) Threshold =20 (e) Threshold =30 (f) Threshold =40

Figure 5.24. SWFI Social Structure after 500 steps using Gauss markov

Table 5.14. Closeness Centrality of SWFI Social Structure using Gauss Markov

time steps	**Threshold (%)**	**Average of Closeness centrality**	
		In Closeness	Out Closeness
500	2	1.617	15.907
	5	1.616	15.543
	10	1.318	1.869
	20	1.013	1.014
	30	1.004	1.004
	40	1.001	1.001
1000	2	1.620	18.249
	5	1.620	18.249
	10	1.616	15.685
	20	1.005	1.005
	30	0.000	0.000
	40	0.000	0.000



(a) Threshold =2 (b) Threshold =3 (c) Threshold =10

(d) Threshold =20 (e) Threshold =30 (f) Threshold =40

Figure 5.25. SWFI Social Structure after 1000 steps using Gauss markov

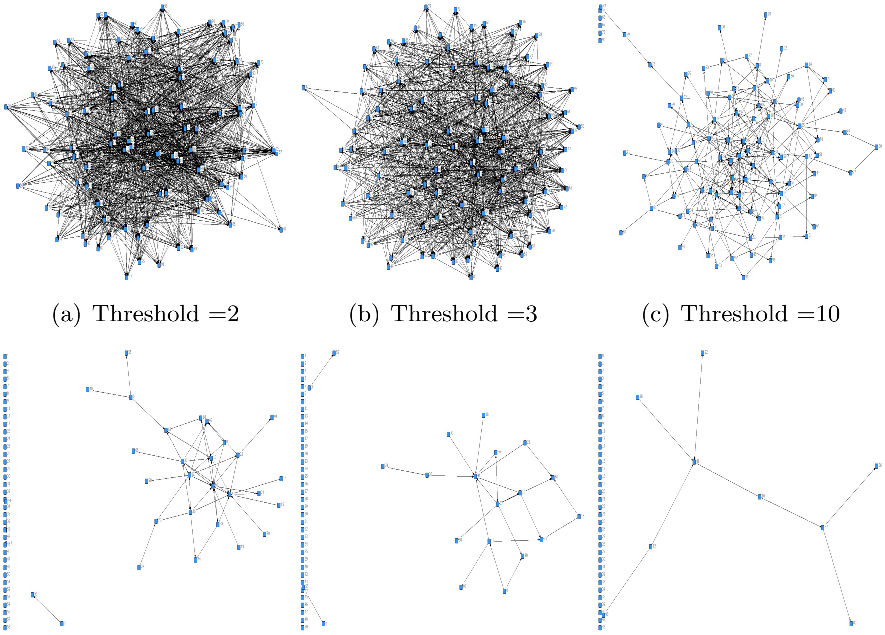
5.7.4 Social Structure using D-GM

In D-GM mobility model, determining the next location for a node is the same as in Guass Markov model. However, instead of a node moving continuously, a node has to stop at certain location. This process indirectly affect the formation of nodes' social structures.

In Figure 5.26 we can observe that all nodes in the network are almost connected with small threshold. This is because nodes have more opportunity to be co-located more frequently. However, when using a big threshold value, it hard for the same nodes to be co-located more frequently in the given sliding window.

A pattern of social structure after 1000 time steps is different than the social structure after 500 time steps. This indicates that the social structure using the SWFI approach captures different social structures based on the current nodes interactions. This is shown in Figures 5.26 and 5.27.

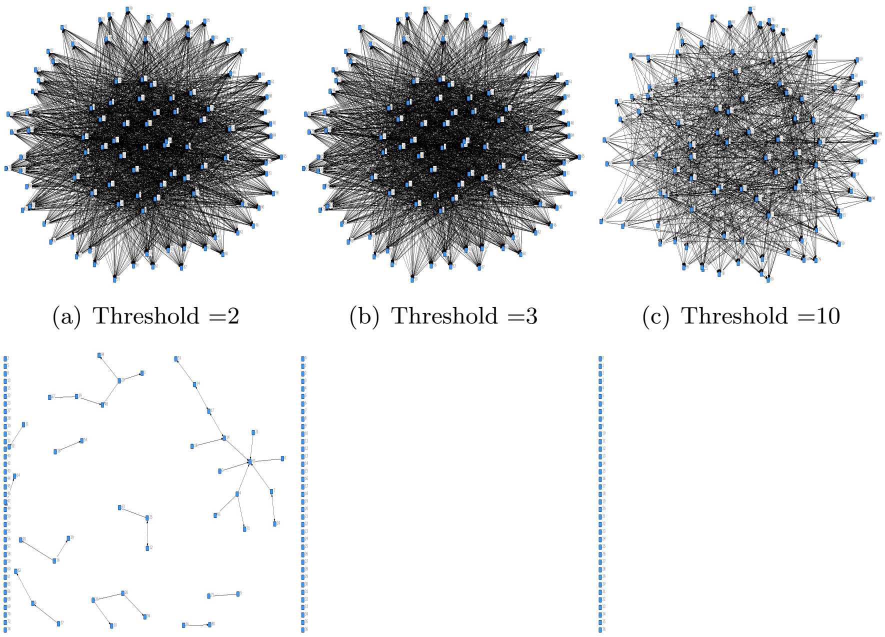
Looking at the statistics in Table 5.15, the average value of closeness centrality is very high. This means that nodes are very close to each other. However, when using a big threshold value, only a few nodes have relationship between them. This indicates that from the D-GM mobility model a close distance between nodes is achieved when using a small threshold value.



(a) Threshold =2 (b) Threshold =3 (c) Threshold =10

(d) Threshold =20 (e) Threshold =30 (f) Threshold =40

Figure 5.26. SWFI Social Structure after 500 steps using DGM



(a) Threshold =2 (b) Threshold =3 (c) Threshold =10

(d) Threshold =20 (e) Threshold =30 (f) Threshold =40

Figure 5.27. SWFI Social Structure after 1000 steps using DGM

Table 5.15. Closeness Centrality of SWFI Social Structure using D-GM

time steps	Threshold (%)	Average of Closeness centrality	
		In Closeness	Out Closeness
500	2	1.620	18.174
	5	1.619	17.619
	10	1.597	8.053
	20	1.022	1.026
	30	1.004	1.005
	40	1.001	1.001
1000	2	1.619	17.630
	5	1.618	16.483
	10	1.615	14.737
	20	1.034	1.037
	30	0.000	0.000
	40	0.000	0.000

5.8 Conclusion

In this chapter three possible ways of forming social structure through the frequency of interaction between nodes has been investigated. The approaches are social structure based on average frequency interactions (section 5.2.1), social structure based on periodicity frequency interactions (section 5.2.2) and social structure based on Sliding Window (section 5.2.3). These represent different ways in which nodes can be configured to define their neighbors.

In general, different mobility models affect the construction of social structures. This is due to the fact that mobility models determine the movement of nodes which indirectly affect the node frequency interactions. Therefore it is essential to chose an appropriate mobility model to have a better social structure.

The average frequency interaction approach captures a general social structure of nodes where the social structure is formulated based on the nodes average interactions. This approach is useful for predicting a general social structure between nodes. From the experiment results, the best social structure score for out closeness centrality measurement is 10 %. This score indicates that this approach is not a good enough to be used as a guidance for neighbor selection in opportunistic networks.

The periodicity interaction frequency approach is focus on capturing the social relationships between nodes in a given period of time. This approach is more focus than the first approach where it able to capture a social structure for a small period of time. Moreover, this approach also provides a flexibility to capture a social structure of a node by simply varying the size of the period of time. Based on the experiment results, the best social structures is achieved under the Gauss Markov mobility model. The best score in terms of out closeness centrality is 14.236 %. This approach is better than the average interaction frequency approach.

The sliding window approach captures a current social relationship between nodes. This approach is more focus than the two approaches in terms of capturing the current node's social structure. As compared to periodicity interaction frequency approach, this approach is more accurate in recording the current nodes' relationships with other nodes because it able to capture social structures of nodes at any given time. Moreover, with the sliding window approach a dynamic change of nodes interactions can be monitored closely. From this chapter perspective, sliding window approach is the most suitable approach to be deployed to capture the social relationships among mobile nodes in opportunistic networks. This approach is also surpasses other two approaches that presented in this chapter in terms of out closeness centrality. Based on the results, the best score of this approach for out closeness centrality is 18.249 %.

From this chapter we learnt that different mobility models and social structure formation technique produce different formation of social structures. This finding is useful for our research because we will use the best social structure approach as neighbour selection to assist information dissemination. We would expect a better performance when social structure captures the dynamic nature of the mobility and interactions. This appears to be supported by the results, since the best social structure formation technique that we discovered uses a Sliding Window (SW) technique. This technique gives a better picture of the nodes social relationship compared to other techniques in terms of relative distance between nodes in the network. Not only that, using this techniques we can capture the node relationship social structure at all points in time. Hence, we use this as the basis for further development in chapter 6.

Chapter 6

INFORMATION SPREADING USING PUSH TECHNIQUES AND SOCIAL STRUCTURE

6.1 Introduction

The purpose of this chapter is to introduce different push techniques and investigate their information dissemination performance and overhead cost. The push techniques that developed here are explained in Section 6.4. We have discovered in chapter 5 that social structures can be defined for various mobility models. We now want to investigate whether these social structures are useful for disseminating information effectively.

In this chapter, we use the mobility models that were introduced in Chapter 4 for investigation. The experimentation results for each push technique with each mobility model is presented in Section 6.6. The performance of each technique is evaluated based on the key performance indicators (KPI) that are described in Section 6.3.

6.2 Using Social Structure

In this section we consider using a social structure to make information dissemination more efficient. Our aim is to investigate the impact of prioritising dissemination through a social structure on information spread performance (i.e, time to receive an artifact). We define a social structure taking into account how frequently nodes interact. Our approach uses a local interaction history to form relationships (i.e edges) between nodes.

For this chapter, we use a sliding window interaction frequency (SWFI) approach which was introduced Chapter 5 in Section 5.2.3. This approach uses a sliding window to determine a node interactions frequency to form a social structure.

6.2.1 Definition of Terms

The following are the terms used in relation to the Social Structure concept.

- Social Structure List (SSL) - is a list of other nodes that are identified as those that are frequently seen by a particular node.
- Social Structure Quota (SSQ) - the maximum number of nodes permitted in (SSL).
- Threshold - the minimum frequency of a node found in SW to be included in SSL.
- Link - is an edge between nodes and represents nodes meeting each other more than the threshold value.
- Total Meeting Frequency (TMF) - is the total number of a particular node detected in a sliding window.
- Interaction History List (IHL) - this is a record that shows a list of nodes that have received information from the current node.

6.2.2 Information Forwarding using Social Structure

We experiment in using the social structure defined in 6.2 and assume that nodes always forward content to a node with whom they have a social link. Node selection is needed when a node has to decide who to push to. In the basic flooding algorithm, information is pushed at every meeting. This scenario contributes to significant overhead costs. With a social structure approach information can be channeled through a population i.e. using relationships between the nodes that frequently meet. For example, let S be a set of nodes that are in range with node p and let Q be a set of nodes that are listed in p's SSL. For example, let us take

$S = \{a, b, c, d, e, f, g\}$

$Q = \{a, e, g\}$

From the SSL, node p is able to direct information to nodes that it has established a relationship with either a,e or g. In turn, these nodes can forward to a node in their SSL. In this way flooding is controlled.

In our approach, at every time step, the SSL of a node is updated. Algorithm 6 shows the SSL update process. Noted that, the SSL is modified when a new node is found in the SW that has TMF bigger than the threshold value.

Algorithm 6 SSL update process

```
Update the SW
Scan the SW and calculate the TMF for each individual node that is detected in the
current SW.
for i = 1 to size of SW do
  curNode = slot[i]
  if curNode found in SSL then
    get the curNode TMF
    if curNode's TMF < threshold then
      remove curNode from SSL
    end if
  else
    get the curNode TMF
    if curNode's TMF > threshold then
      if SSL is not full then
        Add curNode into SSL
      end if
    end if
  end if
end for
```

6.2.3 General Assumptions on Social Structure

The following assumptions are used in our study.

- The information is homogenous and all nodes want the same information.
- The information originates at the information source and can be relayed by any node.
- Each node has a fixed SSQ and SW.
- A link is removed when the frequency of a node found in the SW is less than the threshold value.
- A connection needs to be established prior to push an artifact.

- In the case that more than one node that is in range is in a node's SSL, a random selection between them is performed when choosing who to push or interact with.

6.3 Key Performance Indicators

In order to measure the performance of dissemination across a social structure, we introduce two metrics. They are *Information Profile* and *Average Overhead Push Cost*. Information Profile is used to measure the speed of information dissemination. This is can be observed through the shape of information distribution that is how many node have received an artifact at a particular time. The Average Overhead Push Cost is used to measure how many individual forwarding processes occur to spread information to all nodes.

6.3.1 Information profile

The number of nodes that have received information are recorded at every time step *cumulatively*. At the end of the simulation, the average over a number of runs of the accumulative number of nodes that possess information is calculated. This graph is plotted to investigate the information profile. The purpose of this KPI is to measure how quickly the information (from a single source) is disseminated to all nodes. We also provide addition analysis using *Marginal Information Profile (MIP) of time*. This analysis investigates the percentage of nodes with the artifact for every unit of time. Equation 6.3.1 shows how the MIP of time is measured.

$$MIP_t = PNA_t - PNA_{t-1} \tag{6.3.1}$$

In Equation 6.3.1, PNA_t is the percentage of nodes with the artifact at time t which measures the number of nodes that discover artifact at the given time t. The change in PNA can be discovered through Equation 6.3.1.

6.3.2 Average Push Overhead costs

Push cost is the number of pushes a node produces. A push is when node X forwards information to node Y. This cost is measured *cumulatively.* At the end of the simulation, the average of total number of costs for each trial is calculated. This metric indicates how many pushes involved in each experiment. Furthermore, we provide an additional analysis on *Marginal Overhead Cost (MOC) of time.* This analysis investigates the rate of change of push overhead costs when one unit of simulation time is increased. Equation 6.3.2 shows how the MOC is measured.

$$MOC_t = TOC_t - TOC_{t-1} \qquad (6.3.2)$$

In Equation 6.3.2 TOC_t is the Total of overhead cost at time t. We use the minimal MOC which is denoted as T_{MOC}. If $MOC_t < T_{MOC}$, then the new value of T_{MOC} is MOC_t otherwise the T_{MOC} remains the same.

6.4 Different types of Push Techniques

In this section we provide details of each push technique that is use in this chapter. In general, we can classify the push approaches into two groups: a *push technique with social structure* and a *push technique without structure.* The push techniques with structure use a social structure to select which node to interact with, whereas the other technique (push without structure) chooses its peer randomly. The social structure represents those nodes with which familiarity is maintained. There are five different push techniques introduced in the following sections. These are; *Push with Structure (Section 6.4.1)*; *Push Probability with Structure (Section 6.4.2)*; *Push Probability without Structure (Section 6.4.3)*; *Push Once without Structure (Section 6.4.4)*, and *Push Once with Structure (Section 6.4.5).* All of these push techniques are triggered when there is an interaction between nodes.

6.4.1 Push with Structure (PWS)

The Push with Structure approach is the same as the basic flooding technique. But instead of pushing information to any arbitrary node, a node only pushes to nodes that is listed in the node's SSL. Therefore even though a node is in range of other nodes, no interaction is initiated if none of them are listed in SSL of a current node. Algorithm 7 describes the Push with Structure processes.

Algorithm 7 Push with Structure for node x

```
if node y is in SSL of node x then
    if node x has information then
        x push information to node y
    end if
end if
```

6.4.2 Push Probability with Structure (PPWS)

We assume that each node has the same probability default value for pushing information to others. So, when both nodes are in range and if uniform a random selection value is higher that the default probability, then the node pushes information to its peer. Algorithm 8 shows the Push probability with structure algorithm.

Algorithm 8 Push Probability with Structure for node x

```
if node y is in SSL of node x then
    rndNum := Random number in [0,1]
    if node x has information then
        if rndNum < push probability then
            node x push information to node y
        end if
    end if
end if
```

6.4.3 Push Probability without Structure (PPWOS)

Push Probability without Structure is derived from Algorithm 2 (Pure Push) in chapter 4. In this algorithm the Push Probability without Structure pushes information based on the probability. A node does not simply push information to any nodes but it works based on a probabilistic manner. Algorithm 9 presents the detail of Push Probability without Structure processes.

Algorithm 9 Push Probability without Structure for node x

```
rndNum := Random number in [0,1]
if node x has information then
    if rndNum < push probability then
        node x push information to node y
    end if
end if
```

6.4.4 Push Once without Structure (POWOS)

Push Once without Structure is a slightly more intelligent push approach. A node only pushes information to nodes that it has not pushed to before. This approach assumes that each node has enough memory space to keep records of past pushing events. The records is important because it helps a node to decide to push information or not. The records are maintained in a Interaction History List (IHL). Algorithm 10 shows the process of Push Once without Structure approach.

Algorithm 10 Push Once without Structure for node x

```
if node x has information then
    if node y is not in IHL of node x then
        node x push information to node y
        add node y into IHL of node x
    end if
end if
```

6.4.5 Push Once with Structure (POWS)

The Push Once with Structure algorithm uses a social structure to push information. The social structure is used to identify nodes to interact with, whereas the Push Once without Structure chooses node randomly. Algorithm 11 shows the processes of Push Once with Structure.

Algorithm 11 Push Once with Structure for node x

```
if node x has information then
    if node y is in SSL of node x then
        if node y is not in IHL of node x then
            node x push information to node y
            add node y into IHL of node x
        end if
    end if
end if
```

6.4.6 Overheads

Besides the push and query overheads, all the techniques that use the social structure approach to disseminate information incur extra overheads. These overheads are caused by the process of storing, maintaining and updating the following elements:

- Interaction History List (IHL)
- Social Structure List (SSL)
- Social Structure Quota (SSQ)
- Total Meeting List (TMF)

Note that, these processes are performed locally, so we consider these as internal overheads, and they are not addressed in this thesis.

6.5 Experimentation

The aim of the experiments is to investigate whether is it possible to minimize the overhead costs in information spreading while trying to maintain the performance as close as possible to the flooding technique which is best performing due to is greedy nature. Table 6.1 displays the experimental parameter settings that are used in this chapter.

Table 6.1. Parameters Setting

Parameters	**Setting**
Mobility Models	Random Walk, Random Waypoint, Gauss Markov, D-GM
Size of nodes	100 nodes
Number of information source	1
Information Source coordinate	x = 250m , y = 250m
Size of simulation plane	500m x 500m
Simulation duration	15 minutes (9000 time steps)
Number of trials	50 times
Social structure quota	10 members
Window Size	40

Because different mobility models have different impact on information spreading [25], we have also used different mobility models for each push technique introduced in this chapter. The following results are organized based on different push techniques.

6.6 Results

To recap, there are five different push approaches introduced in this chapter. There are **Push with structure**, **Push probability with structure**, **Push probability without structure**, **Push Once with structure** and **Push Once without Structure**. Each of this approach is analyzed with different mobility models (Random Walk, Random Waypoint, Gauss Markov, and D-GM). The following results are organized based on different push approaches.

6.6.1 Result - Push with Structure

6.6.1.1 Push with Structure using Random Walk

In this section, a threshold value is a minimum frequency that a particular node found in SW to be included in SSL (Social Structure List). In Figure 6.1, we observe that increasing the threshold value reduces the number of nodes that receive an artifact over time. This is due to the fact that with a high threshold value, there is more constraint on establishing an interaction. This affects the number of nodes that can be included in SSL. Therefore, the opportunity of nodes pushing information is decreases as the threshold value is increased. This is why in Figure 6.1, the social structure with threshold=40 takes a longer time to make the information available to all nodes in the networks.

In addition, the marginal analysis in Figure 6.2 confirms that lesser constraints make the information dissemination accelerate at the early stage. This is due to the fact that the node has an opportunity to push information to different nodes at the early stage as nodes easy to establish relationship with other nodes (i.e threshold=2). However, with a high constraint (i.e threshold=40) the marginal information profile of time shows a small change in the percentage of nodes with an artifact when a unit of time is increased. This is indicates that a node needs more time in order to establish more relationships with other nodes. Consequently, it delays the information dissemination to all nodes in the networks.

As we can observe from the Table 6.2 and Figure 6.3, the threshold influences the frequency of interaction and the formation of social structure. Note that the links of the social structure in Figure 6.3 is a snapshot of 1000 time steps simulation. We can see

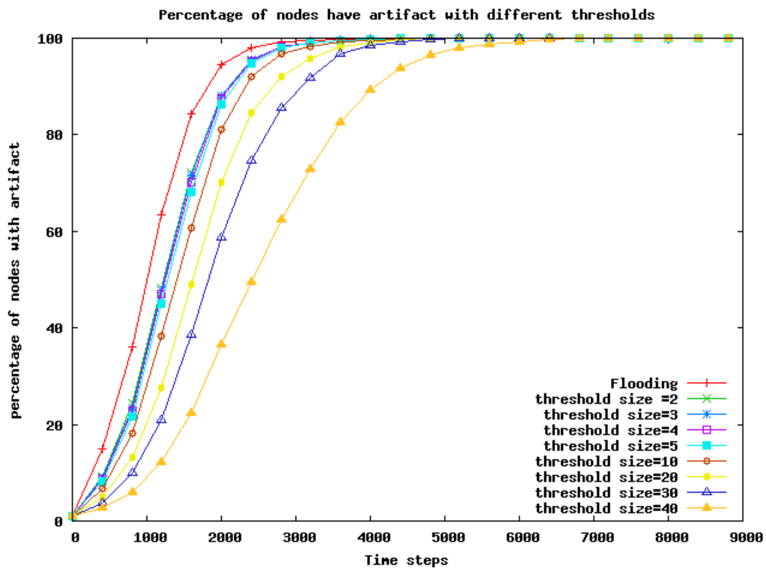


Figure 6.1. Information Profile for Push with structure with different Threshold value using Random Walk

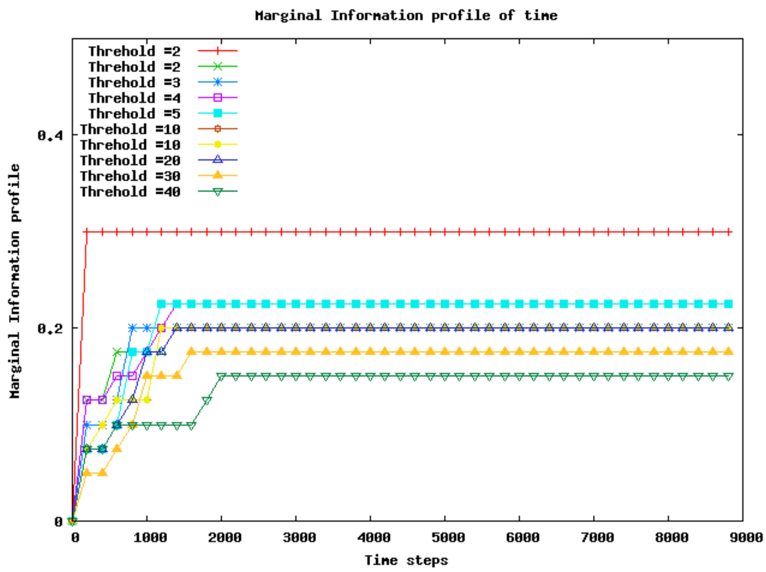


Figure 6.2. Marginal Information profile of time using Random Walk

that with a small value of threshold, more links is formed as it is easy to maintain a small interaction frequency between nodes. This helps to boost the information dissemination performance. With a strictness (high threshold value) of social structure formation, it will limit the opportunity of node to push information to other nodes. As a result, information spreads very slowly when more constraint (high threshold) is imposed.

The overhead cost (number of pushes) is dependent on the frequency of interaction.

Table 6.2. Average node interaction using Random Walk

Threshold	Average interactions per node
2	26.762
3	26.555
4	26.361
5	26.166
10	25.163
20	22.526
30	16.252
40	10.611

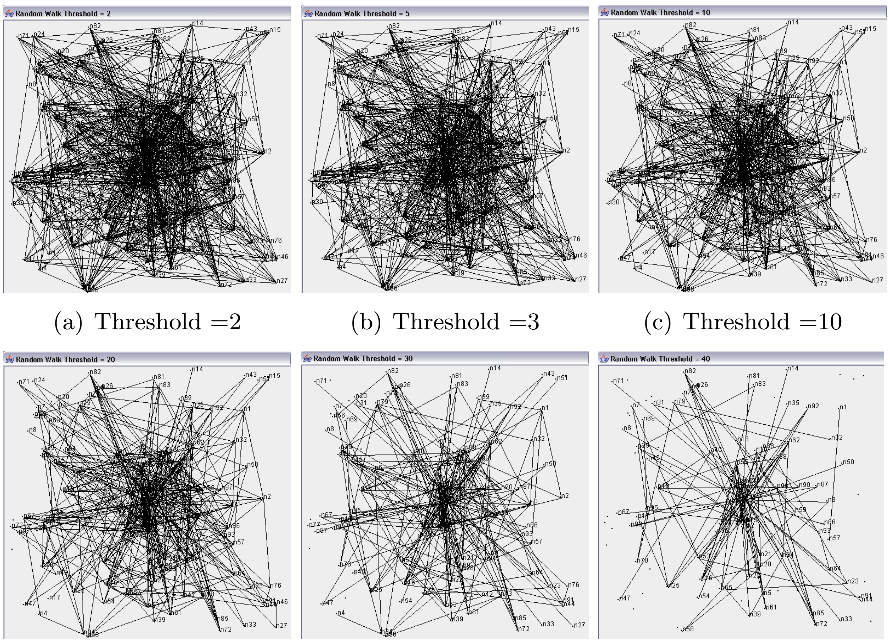
The more chance of node to interact with other nodes, the more likely the overhead cost is increased. Looking at the overhead costs in Figure 6.4, increasing the threshold value reduces the overhead costs. This is because not many nodes are involved in pushing information as nodes are restricted to push to a particular node only (i.e. in node's SSL).

Looking at the marginal overhead cost in Figure 6.5, the overhead cost increases according to the threshold value setting. When the threshold value is small, the change of overhead costs over time increases rapidly. This is because more nodes find it easy to establish a connection with other nodes as the number of nodes with the artifact increases rapidly over the time.

Figure 6.6 shows the relationship between the overhead costs and the percentage of information dissemination. As we can see from the figure, increasing the overhead cost also increases the percentage of nodes with the information artifact. This indicates that information dissemination has direct impact on the cost. Note that no time reference is considered in Figure 6.6. Looking at Figure 6.6, at 70% to 100% of nodes with an artifact, the overhead costs in the social structure approach is far better than flooding. Further more, the high constraint (i.e threshold = 40) shows a better combination between overhead cost and number of nodes with an artifact. This is due to the fact that the social structure approach with a high threshold has a better cost utilization but it has high delay on disseminating information to all nodes in the network.

6.6.1.2 Push with Structure using Random Waypoint

The result in Figure 6.7 has the same trend as Figure 6.1. The information dissemination performance is decreasing when the threshold value is increased. The threshold value is



(a) Threshold =2 (b) Threshold =3 (c) Threshold =10

(d) Threshold =20 (e) Threshold =30 (f) Threshold =40

Figure 6.3. Push with structure technique using Random Walk with different Threshold

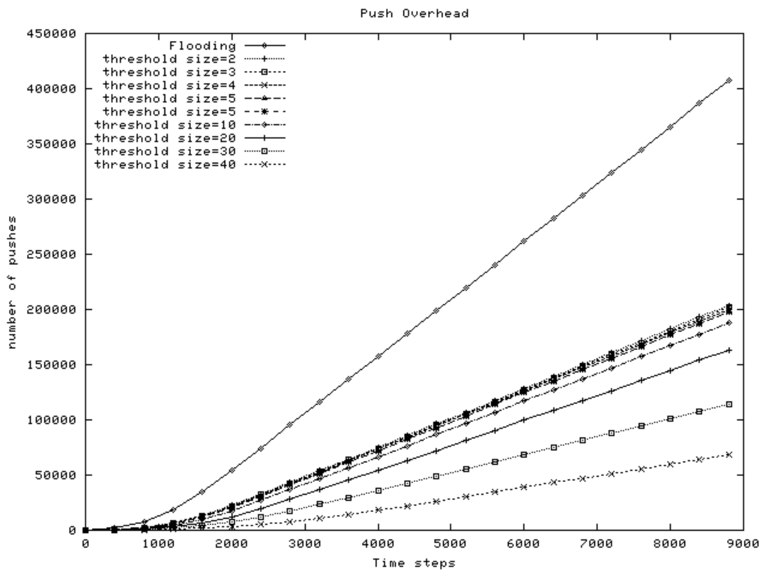


Figure 6.4. Average Push Overhead Costs for Push with structure technique using Random Walk

the minimum number of frequency of a particular nodes found in SW to be included in SSL. This is due to the fact that, increasing the threshold value also means increasing the strictness of social structure link formation. The performance in Figure 6.1 is better than Figure 6.7 because in the Random Walk Mobility model, the nodes have more chance to meet different nodes as compared to Random Waypoint mobility model.

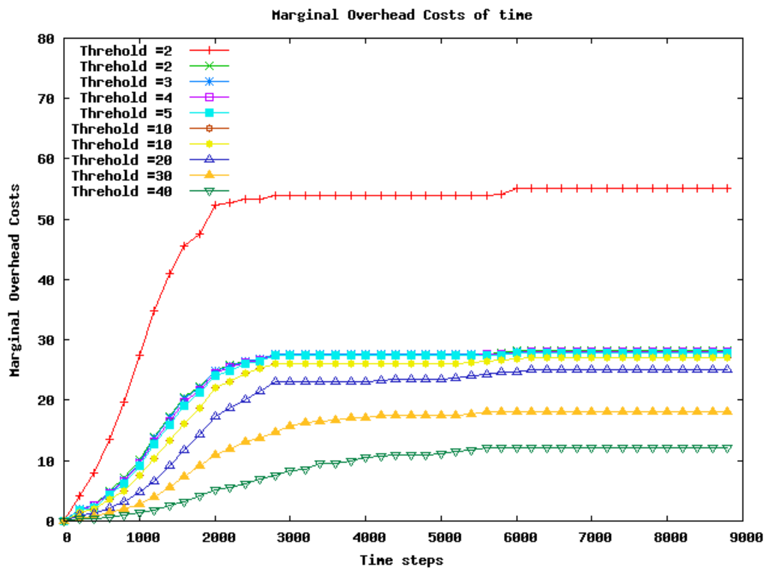


Figure 6.5. Marginal Overhead Costs of time for Push with structure using Random Walk

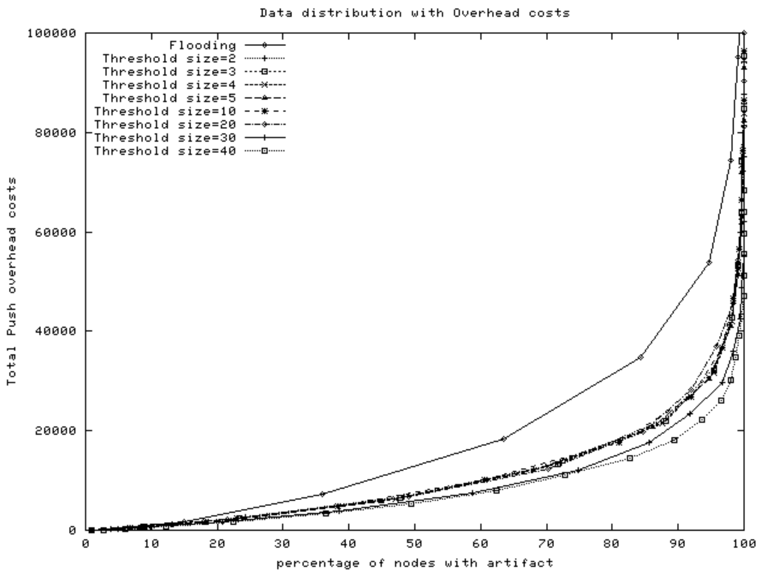


Figure 6.6. Overhead Cost vs Node with artifact for Push with structure using Random Walk

From Figure 6.8, the change in percentage of nodes with an artifact when one more unit of time increases is very small. This is because using the Random Waypoint mobility model, nodes have to stop at certain locations for a period of time before they can move to another location. The stopping attribute reduces the interaction of nodes. This results in a small change in the marginal information profile in Figure 6.8.

The overhead costs in using the Random Waypoint mobility model increases when more

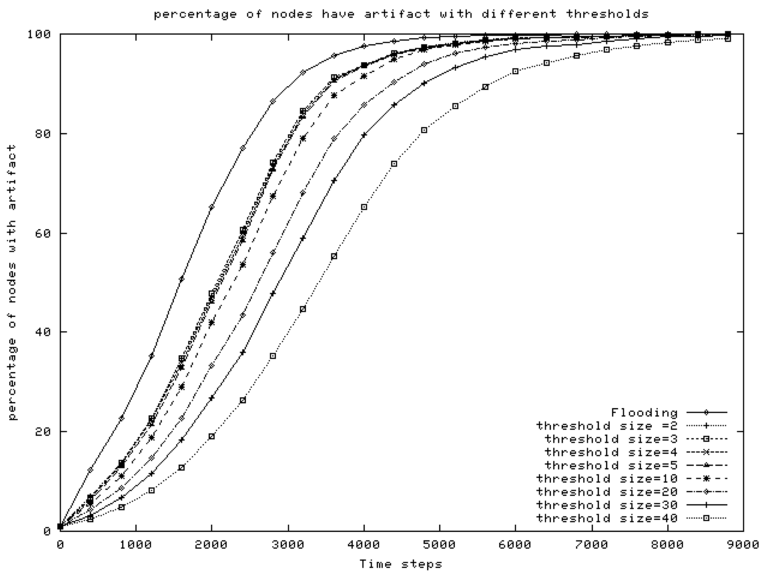


Figure 6.7. Information Profile for Push with structure with different Threshold value using Random Waypoint

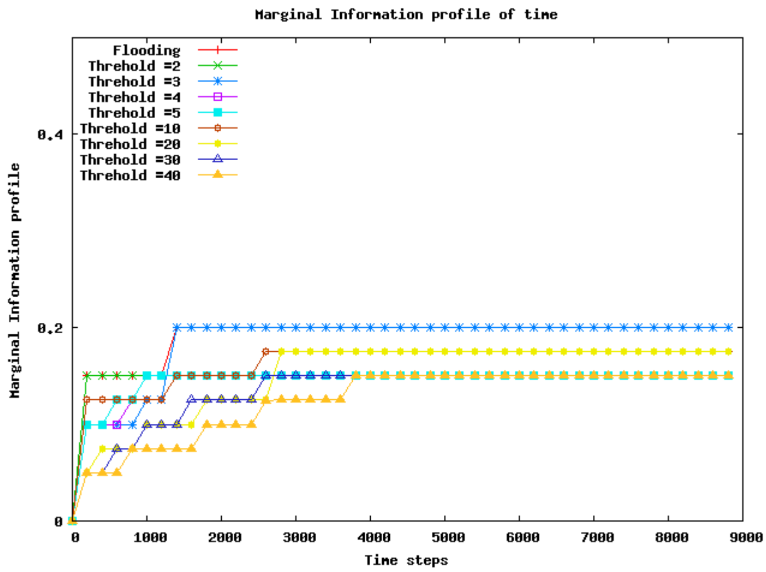


Figure 6.8. Marginal Information profile of time using Random Waypoint

nodes are actively involved in interactions. This is shown in Figure 6.9. This happens because the number of nodes that have an artifact is increased over time, which also increases the number of nodes involved in pushing information. Looking at the marginal overhead costs in Figure 6.10, the change of overhead costs accelerate quickly at the early stage (i.e before t=4000). This is because more nodes are starting to be involved in pushing information. Note that the different acceleration rates in the early stages is because of the strictness of link formation as shown in Figure 6.11.

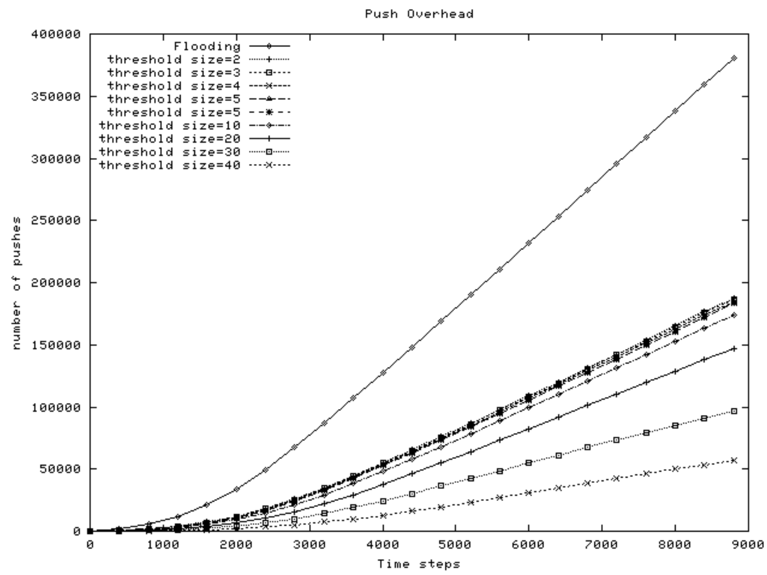


Figure 6.9. Average Push Overhead Costs for Push with structure using Random Waypoint

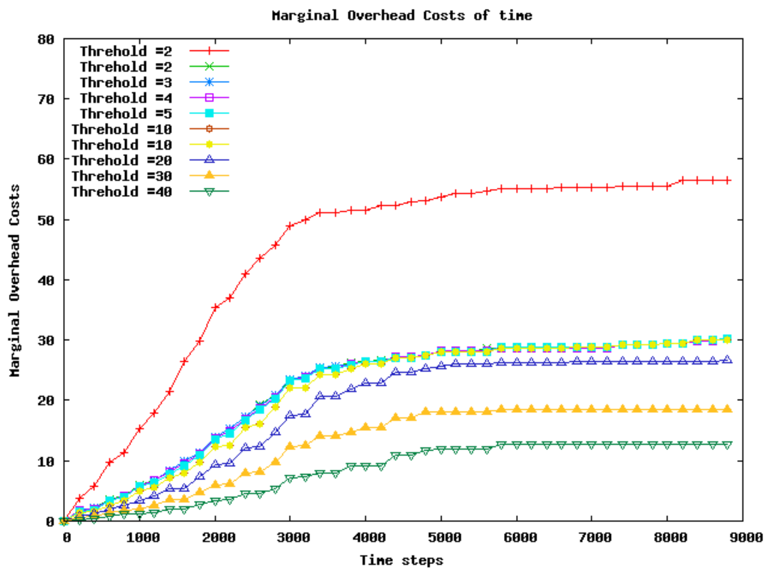


Figure 6.10. Marginal Overhead Costs of time using Random Waypoint

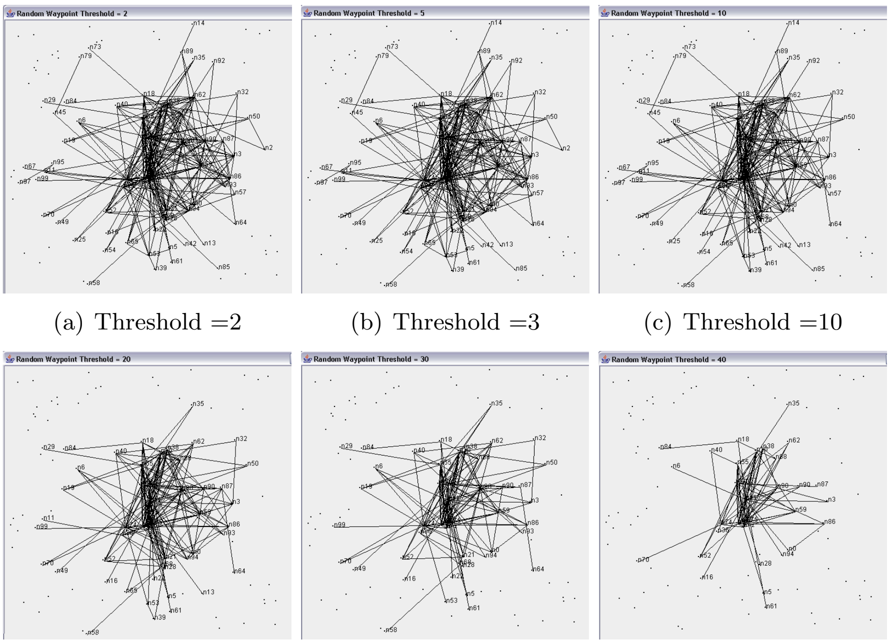
From Table 6.3, we can observe that a high constraint (threshold=40) has a lower average interaction per node. This is because it is difficult to form a social structure with a high constraint as nodes have to be co-located overall for a long consecutive periods of time. As a result, the number of nodes that involved in interactions are decreased when the threshold is increased.

Figure 6.11 shows the social structure links between nodes with different threshold

Table 6.3. Average node interaction using Random Waypoint

Threshold	Average interactions per node
2	27.269
3	27.109
4	26.940
5	26.772
10	25.920
20	23.090
30	16.3112
40	10.760

values. From the figure, we can observe that the density of links decreases as the threshold value is increased. This is because a threshold represent the strictness of links formation. It means that, if the threshold is small (i.e. threshold=2), a link can be formed easily between nodes because the nodes just have to be co-located two times in a SW. However, when the threshold is increased, nodes have to be co-located more frequently (threshold =40) within SW in order to establish a link.



(a) Threshold =2 (b) Threshold =3 (c) Threshold =10

(d) Threshold =20 (e) Threshold =30 (f) Threshold =40

Figure 6.11. Push with structure using Random Waypoint with Differen Threshold

Figure 6.12 shows the relationship between overhead costs and percentage of nodes with artifact. Looking at the figure, as we can expect, flooding has more overhead costs compared to the social structure approach. This is because, in flooding every node is pushing information at every opportunity. Alternatively, in the social structure approach, the information pushing is controlled by the links between nodes. The nodes in social structure only push information when a link between them are available. In another perspective, flooding utilizes any meeting opportunity to maximize the information dissemination whereas social structure approach utilizes social links to minimize the overhead costs.

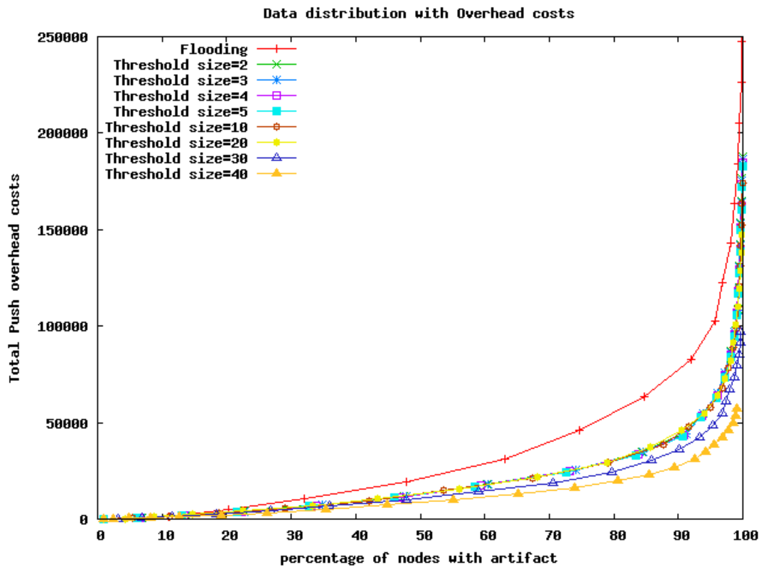


Figure 6.12. Overhead Cost vs Node with artifact using Random Waypoint

6.6.1.3 Push with Structure using Gauss Markov

From Figure 6.13, we can observe that with a small threshold (i.e. between 2-10), the number of nodes which possess an artifact using the social structure approach is close to the number of nodes with an artifact using the flooding approach. However, when the threshold value is further increased, the numbers of nodes that have information are decreased dramatically over the time. This is because with a high threshold it is hard for nodes to be co-located with their social neighbors that are listed in the nodes' social structure. Besides that, nodes moving using the Gauss Markov mobility model changes their direction at every time step, so it is hard for the same nodes to be co-located continuously.

Analyzing the change of number of nodes with the artifact when one unit of time is increased, we observe that the social structure with high threshold value (threshold=40) has very small changes in the number of nodes with the artifact and remain level under

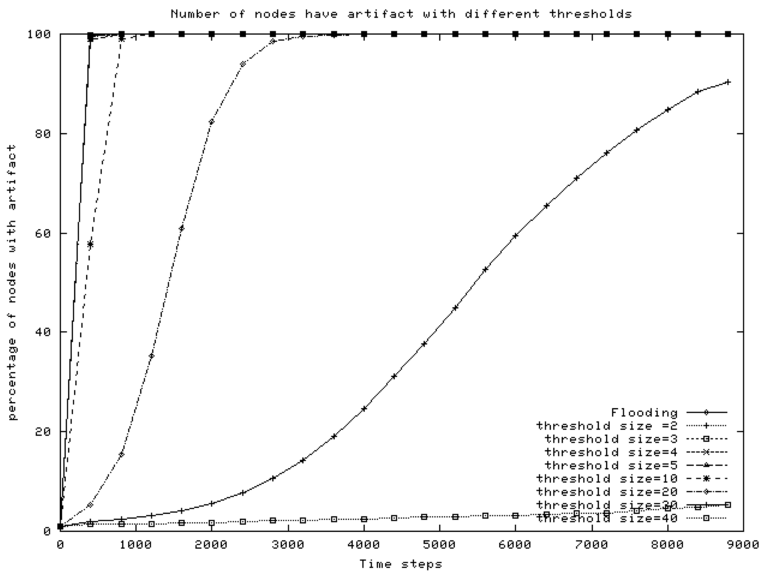


Figure 6.13. Information Profile for Push with structure with different Threshold value using Gauss Markov

0.2% Marginal Information Profile (MIP). This is shows that no further interactions will manage to improve the number of nodes that posses an artifact. But with a small threshold value, as we can see from Figure 6.14, the percentage of nodes that have information increases very quickly over time in the early stage. This is because with a low threshold it is easy for nodes to establish a relationship which results in more received information.

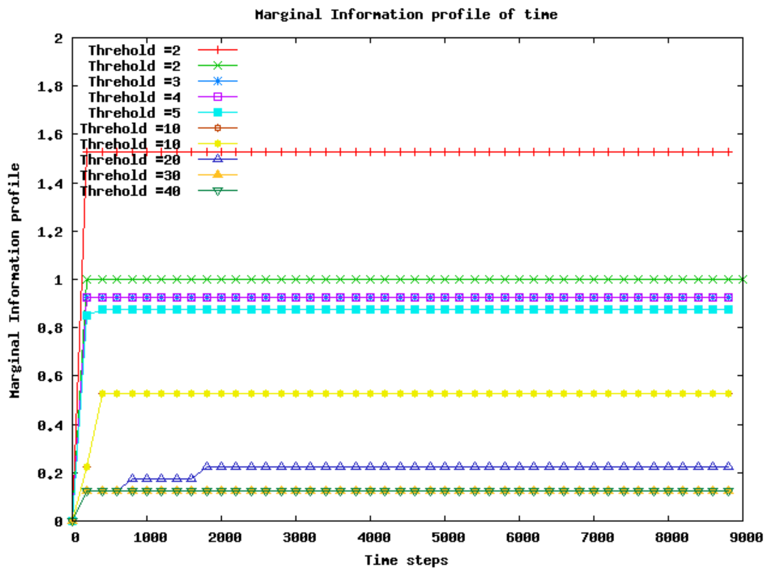


Figure 6.14. Marginal Information profile of time using Gauss Markov

As in the Random Walk and the Random Waypoint results, the overhead costs reduce as the threshold value is increased (Figure 6.15). This is because with a high threshold (i.e. threshold=40), it is difficult to form social links between nodes. This is limiting the number of nodes involved in information pushing, which reduces the overhead costs significantly. Note that this reduction of cost is not cost effective as it has a very low number of nodes with the artifact (refer to Figure 6.13).

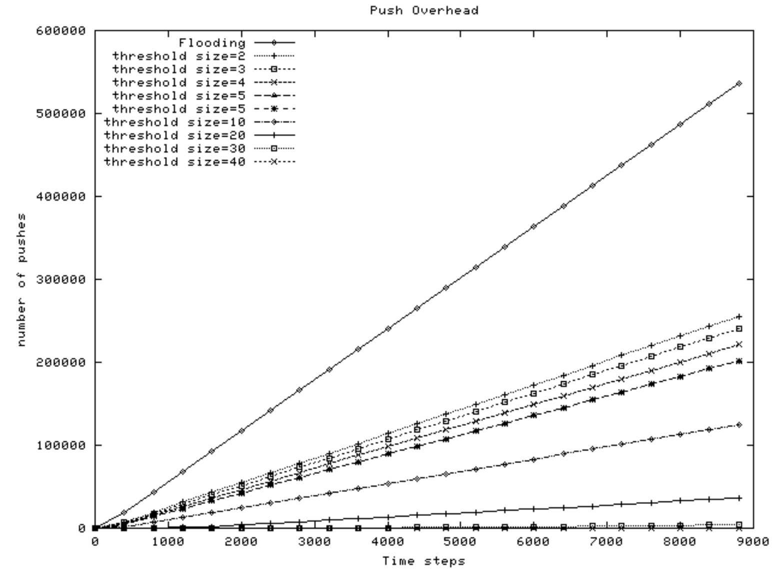


Figure 6.15. Average Push Overhead Costs for Push with structure using Guass Markov

Looking at Figure 6.16, the overhead costs accelerate very quickly in the early stage with the social structure approaches that have small threshold values. This is because a large percentage of nodes received information very quickly at the beginning of the simulation. As more nodes are involved in information pushing which contributes to a large change in overhead costs when a unit of time is changed. The marginal change of overhead cost is reduced as the threshold value is increased. This is due to the fact that not many nodes are able to maintain a link as the nodes have to be co-located frequently.

As we can see from Table 6.4, there is a huge gap in average interaction between a low threshold (threshold=2) and high threshold (threshold=40). This is because the formation of links in the social structure depends on how strict the link formation is. When the threshold is high (threshold=40), it means that nodes have to be co-located continuously within 40 time steps in order to establish a link. With the Gauss Markov mobility model, it is very rare to have the same nodes co-located consecutively for a long period of time because the Gauss Markov mobility model assign different directions at

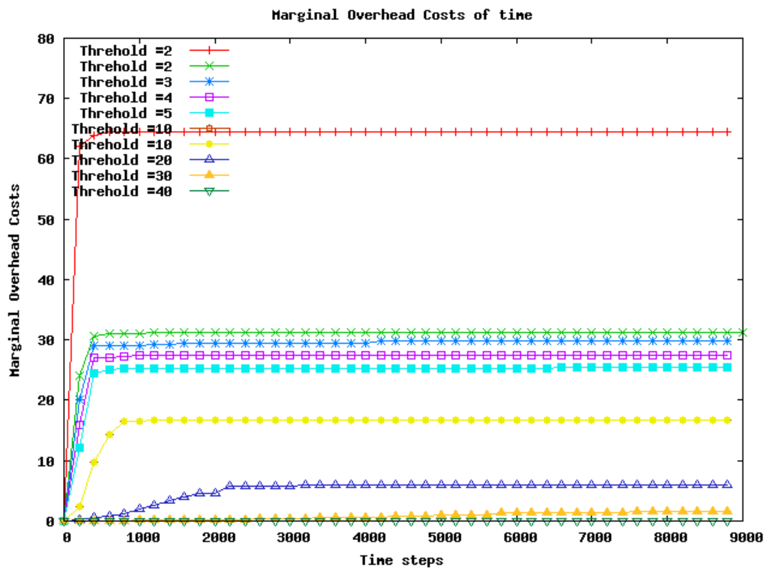


Figure 6.16. Marginal Overhead Costs of time using Gauss Markov

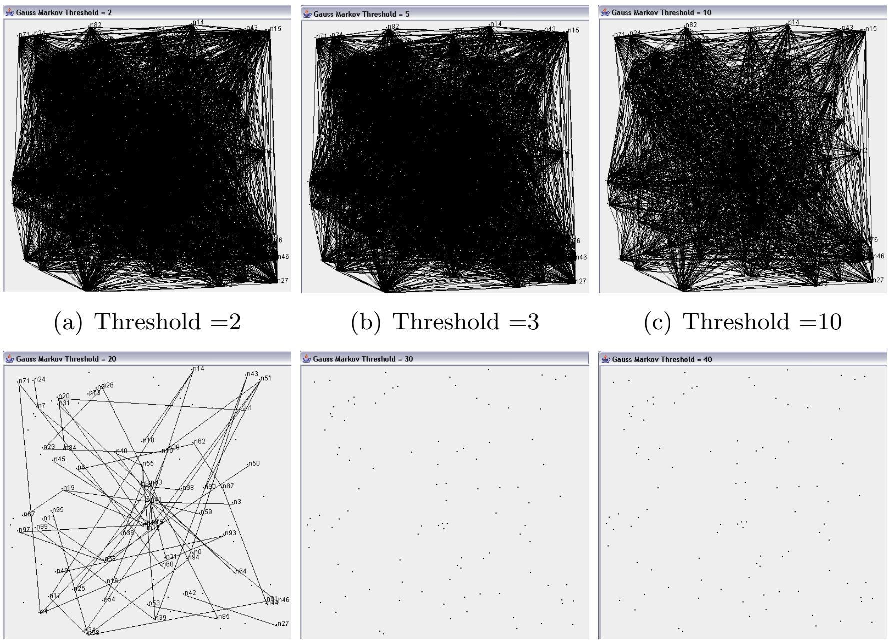
Table 6.4. Average node interaction using Gauss Markov

Threshold	Average interactions per node
2	29.479
3	27.846
4	25.628
5	23.413
10	14.804
20	4.901
30	1.052
40	0.112

every time step.

Besides that, nodes moving using the Gauss Markov mobility model changes its direction at every time step, so it is hard for the same nodes to be co-located continuously.

Figure 6.17 shows the number of links between nodes with a different threshold setting value. Using Gauss Markov Mobility model, there is a high interaction at a lower threshold (between 2- 10). This shows that more nodes in average are co-located consecutively in less that 10 time steps. In addition, this also explains why the average number of interactions (Table 6.4) with the lower threshold is very high. In Table 6.4 and Figure 6.17, we found that no nodes were able to establish connections when the constraint is higher (threshold=40). This is because it is difficult for a node to be co-located with the



(a) Threshold =2 (b) Threshold =3 (c) Threshold =10

(d) Threshold =20 (e) Threshold =30 (f) Threshold =40

Figure 6.17. Push with structure using Gauss Markov with Different Threshold

same node.

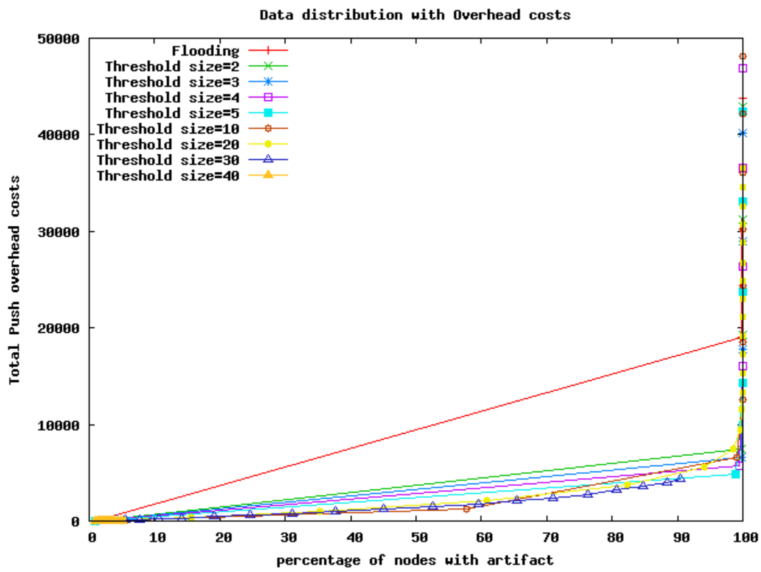


Figure 6.18. Overhead Cost vs Node with artifact using Gauss Markov

Looking at Figure 6.18, effective dissemination of information appears to have a consequence of high overhead costs. The approach with lower overhead costs (threshold=40) results in a very small percentage of nodes with artifact. This is because not many nodes are involved in interactions and consequently a link between them is hard to establish.

Conversely, with the small threshold, it is easy for nodes to establish a link. Therefore there is a lot of interaction overhead costs involved which create a chance for nodes to received an artifact. Note that the figure does not consider time or how quickly the information is available to all nodes. However, this figure is important for understanding the effect of overhead costs on the percentage of nodes with artifact when time is not part of the performance consideration.

6.6.1.4 Push with Structure using D-GM

From Figure 6.19, we can see that the performance of information dissemination of push with structure approach is close to the flooding performance when a threshold is small (i.e. threshold between 2-5). This is because it is easy for nodes to form a social link with each other. As a result, more nodes are infected (possess information) very quickly. As we expected, when the threshold is increased the information dissemination performance is decreased over the time. This is because it is difficult to establish a social link between different nodes.

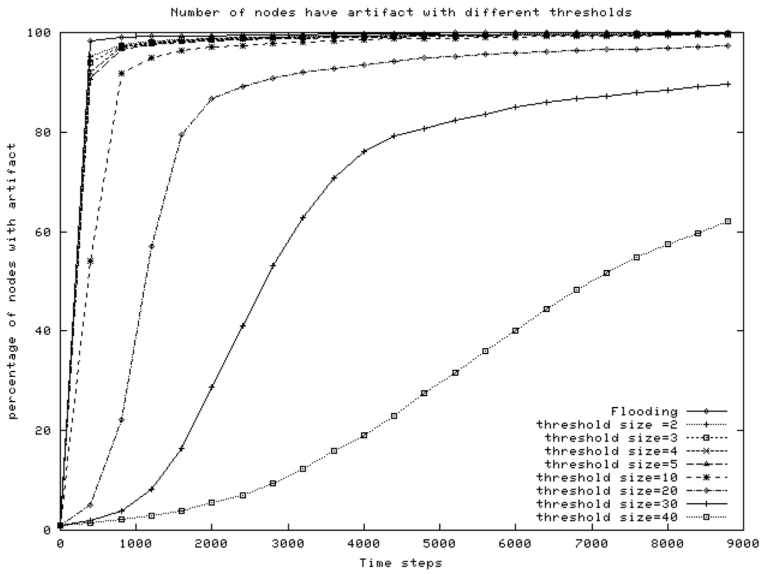


Figure 6.19. Information Profile for Push with structure with different Threshold value using DGM

Looking at the marginal information profile of time, the changes of the number of nodes with an artifact increases drastically over a short time when the threshold is small. This shows that interactions at the early stage helps to discover nodes with the artifact.

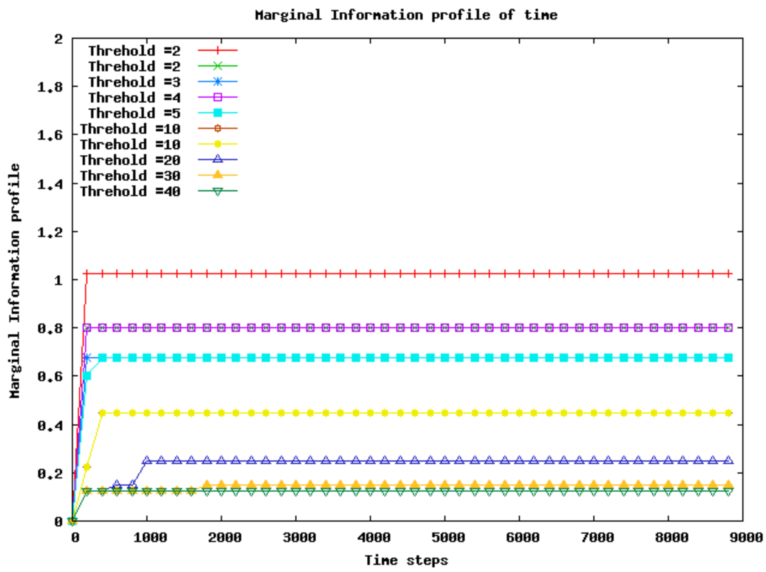


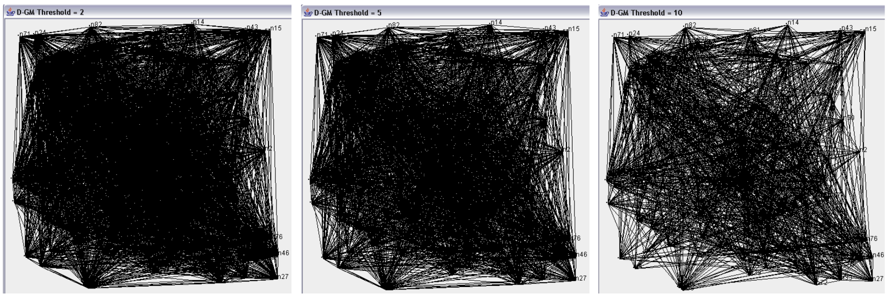
Figure 6.20. Marginal Information profile of time using D-GM

But, with the high threshold, there are very small changes in the information profile when the simulation time is increased by a unit. This is because, at that time many nodes have not discovered an artifact. That Marginal Information Profile is constant throughout the simulation shows that nodes have difficulty in establishing links with different nodes. This decreases the opportunity of information dissemination performance.

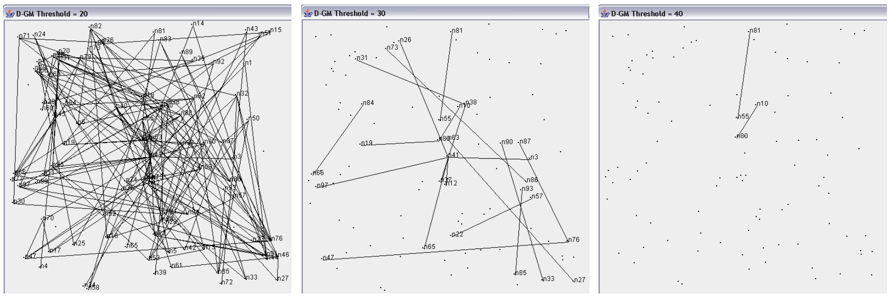
Table 6.5. Average node interaction using D-GM

Threshold	Average interactions per node
2	11.642
3	11.152
4	10.626
5	10.101
10	7.815
20	4.550
30	2.317
40	0.818

In Table 6.5, the average interaction per node is decreased as the threshold value is increased. This is because the constraint of forming a link between nodes is strict as the threshold is increased. In the social structure approach, nodes can only push information (interact) with the nodes that have been co-located more than or equal to the threshold value. This is why a lower threshold in the social structure approach has a high average interaction per node because it is easy for nodes to be co-located consecutively twice rather than forty times.



(a) Threshold =2 (b) Threshold =3 (c) Threshold =10



(d) Threshold =20 (e) Threshold =30 (f) Threshold =40

Figure 6.21. Push with structure using D-GM with Different Threshold

From Figure 6.21 we can observe that there is a high possibility of nodes to interact with each other. This situation helps the information spread to different nodes very quickly. A result of this is that more nodes discover information at the early stage as shown in Figure 6.19. When the threshold is increased the density of links between nodes is significantly reduced. This is because it is difficult to identify neighbors to forward information. Therefore, fewer nodes possess information at the early stage with the high threshold.

Based on the Figure 6.22, there is a big gap in overhead costs between the social structure and flooding approaches. This is because nodes in flooding push information at every opportunity and do not have any restriction in order to push information. However, in the social structure approach, forwarding information is subject to how many links that a node has. As a result, a very small number of nodes involved in forwarding information when a threshold=40 is used. With fewer interactions, the overhead cost also reduced because the overhead cost is actually measures how many forwarding processes occur through out the simulation.

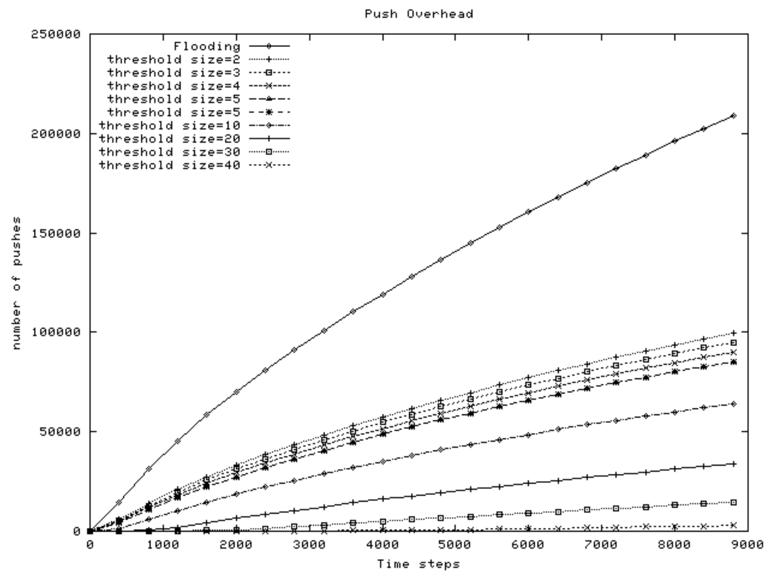


Figure 6.22. Average Push Overhead Costs for Push with structure using D-GM

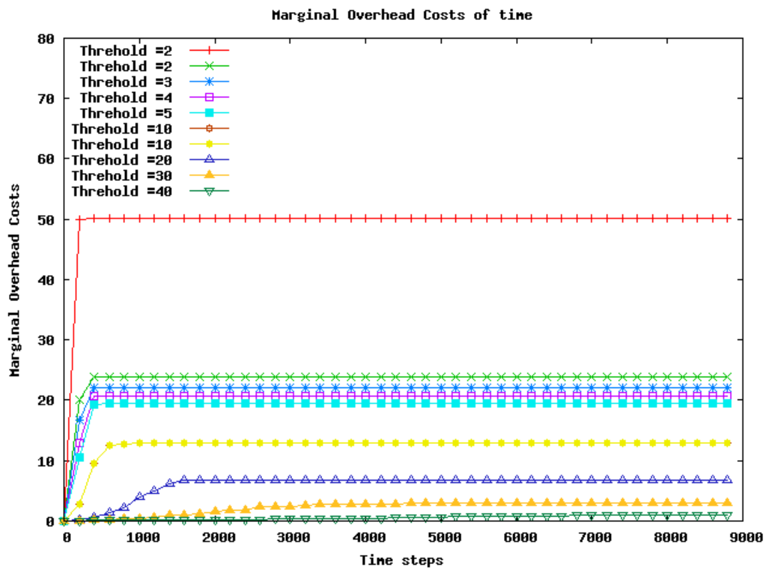


Figure 6.23. Marginal Overhead Costs of time using D-GM

Looking at Figure 6.23, the overhead costs accelerate very quickly at the early stage with the social structure approaches that have small threshold values. This is because a large percentage of nodes receive information very quickly at that time (refer to Figure 6.19). So, more nodes are involved in information pushing which causes the high change in overhead costs when a unit of simulation time is increased. The marginal overhead costs change accordingly with the threshold. This is because the link formation is subject to the threshold strictness. For example the marginal overhead costs for threshold=40 is near to zero. This shows that not many nodes are able to forward information as it is difficult to form a social structure.

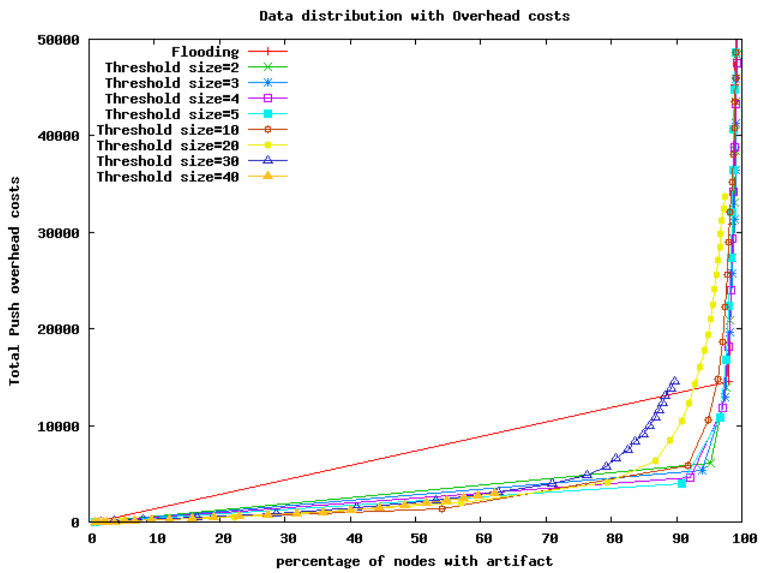


Figure 6.24. Overhead Cost vs Node with artifact using D-GM

From Figure 6.24, we can observe that the approach with a low overhead cost (e.g threshold $\geq$ 40) has a smaller percentage of nodes with an artifact. This is because an interaction between nodes are limited as they have to be co-located more that 40 times consecutively. With the small threshold, it is easy for nodes to establish a link. Therefore more frequent interactions are involved, which helps to disseminate information to all nodes. This condition boosts the overhead costs as more nodes are involved in forwarding information. Note that the figure does not consider the time or how quickly the information is available to all nodes. However, in this figure it is important to understand the effect of overhead costs on the percentage of nodes with artifact when time is not part of the performance consideration.

6.6.2 Results - Push Probability with Structure

The social structure presented hereafter uses a small threshold value (threshold = 2) as the basis for forming the links between nodes. This addition information is added as part of the parameter settings listed in Table 6.1. We use a single threshold value because we want to focus on investigating the impact of push probability with a social structure when different values of probability are applied. Furthermore, we also need to keep the number of experimentation variables are manageable since both push probability and social structure have a large number of parameters.

6.6.2.1 Push Probability with Structure (PPWS) using Random Walk

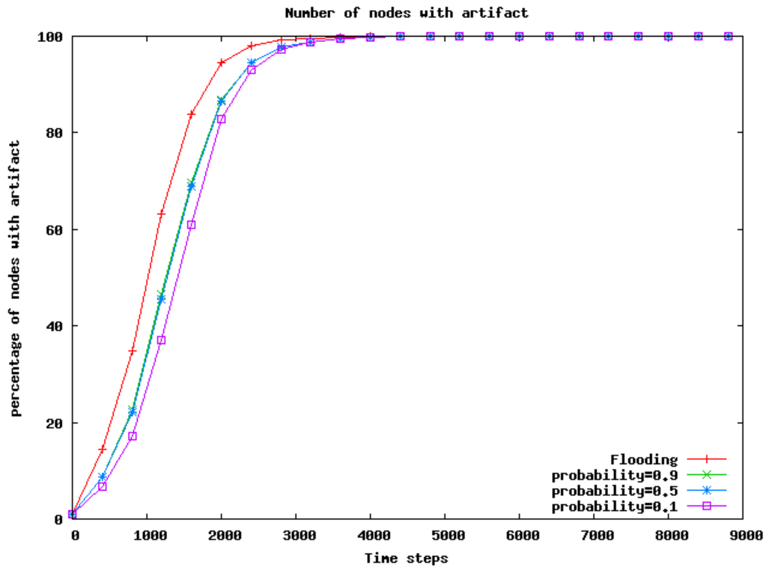


Figure 6.25. Information Profile for Push Probability with Structure using Random Walk

From Figure 6.25, we observe that varying the push probability affects the information dissemination performance. We test a range of different push probability values i.e 0.9, 0.5 and 0.1.

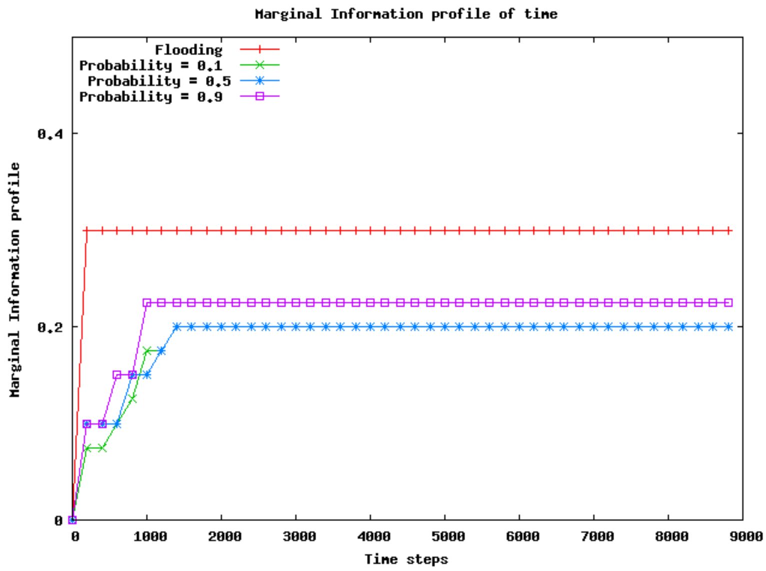


Figure 6.26. Marginal Information profile of time for PPWS using Random walk

Looking at Figure 6.26, the performance increases very quickly when one unit of time is increased at the early stage. This is because not many nodes discover information at the early stage, so the number of nodes that are infected with information has a dramatic effect at the early stage. In the early stage, the MIP is constant for a certain period of time and this happens because of the probability behavior. We can see that the probability with push probability of 0.9 outperforms the lower probability (i.e. 0.1).

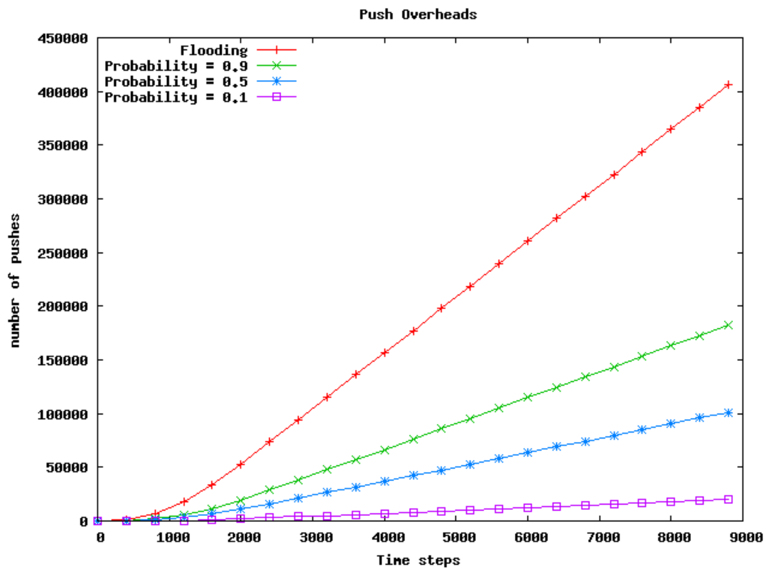


Figure 6.27. Average Push Overhead Costs for PPWS using Random Walk

The overhead costs of the PPWS approach presented in Figure 6.27 shows that the high probability (i.e. probability = 0.9) incurs a high overhead cost. The big gap at a high number of time steps for the PPWS approach is because of the difference in the number of nodes that become involved in pushing the information when different probability value is used. The higher the probability value, the more chance of nodes pushing information to others.

From Figure 6.28, the marginal overhead costs accelerate very quickly at the early stage. This shows that more nodes are actively involved in pushing information when a unit of simulation time is increased. The big gap in marginal overhead cost between flooding and the PPWS with push probability 0.9 is because the PPWS uses a social structure to guide the push process which reduces the number of pushes among nodes. In flooding nodes push information at every meeting which causes the overhead costs to increase.

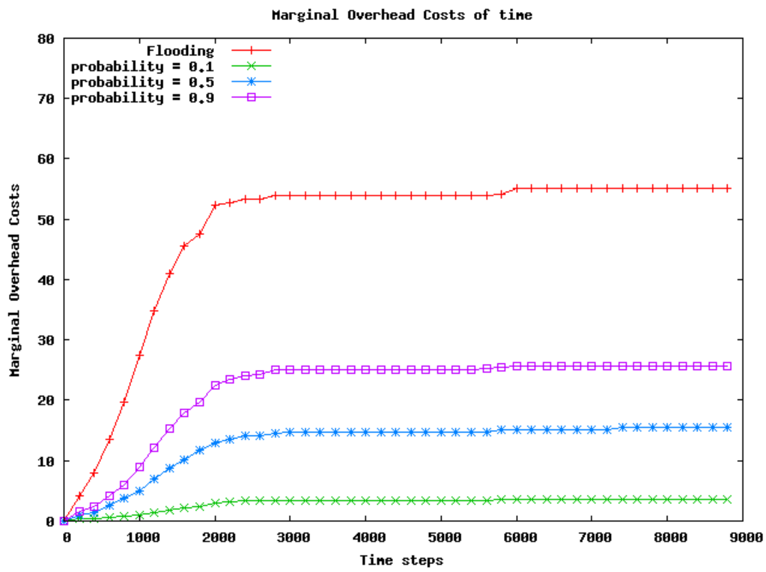


Figure 6.28. Marginal Overhead Costs of time for PPWS using Random Walk

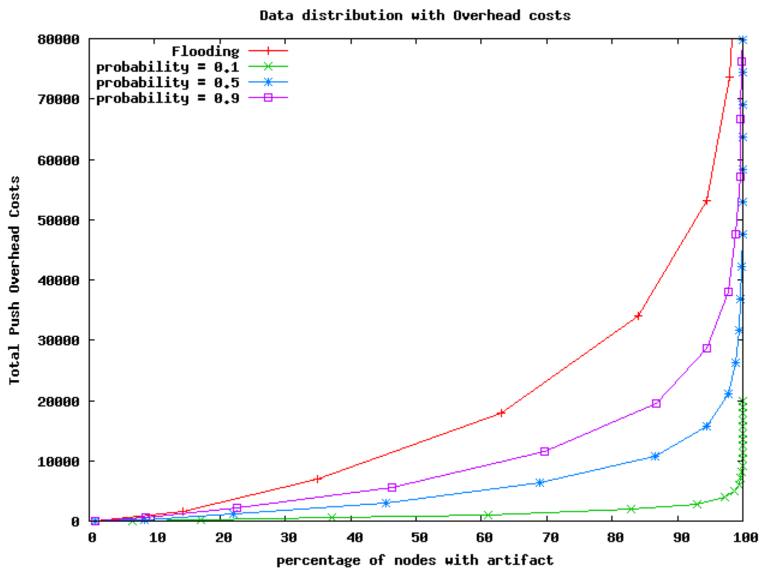


Figure 6.29. Overhead Costs vs Node with artifact for PPWS using Random Walk

Looking at Figure 6.29, the PPWS with probability=0.1 has the lowest overhead costs as compared to other approaches in the graph. However, this does not mean that PPWS with probability=0.1 is the best approach because it takes a longer period of time to disseminate information to all nodes in the network (refer Figure 6.25).

6.6.2.2 Push Probability with Structure using Random Waypoint

The information profiles in Figure 6.30 have the same pattern as in Figure 6.25. A high push probability creates more opportunity for nodes to push information to other nodes. It also shows that different values of push probabilities result in different information profile. This is because the push probability determines the capability of a node to disseminate information to other nodes. Even though a node has information to push, no interactions will be initiated unless the probability value permits.

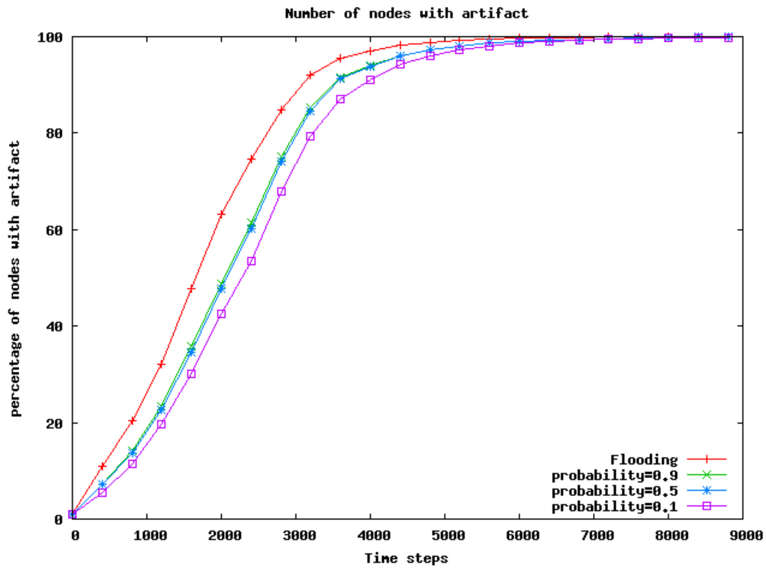


Figure 6.30. Information Profile for PPWS using Random Waypoint

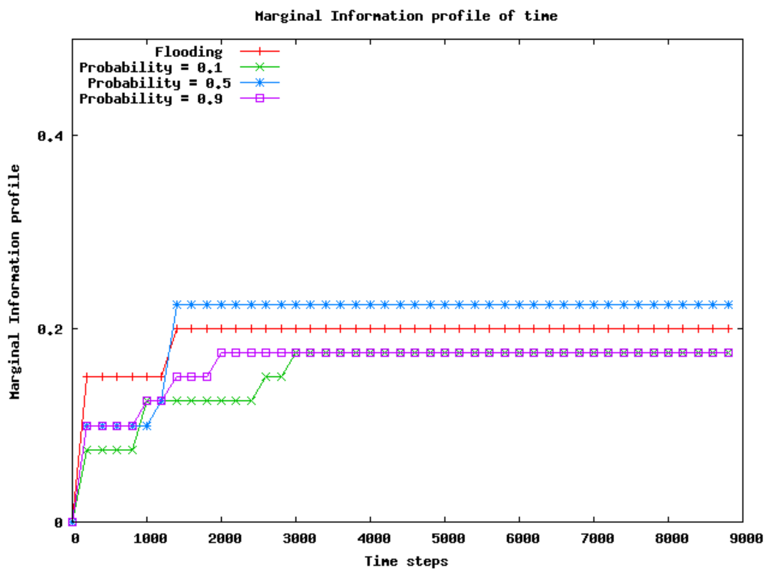


Figure 6.31. Marginal Information profile of time for PPWS using Random Waypoint

Figure 6.31 shows the change of performance (number of nodes with artifact) when one additional unit of time is increased. As can be seen from the figure, the PPWS with push probability=0.5 has marginal information profile higher than flooding after 1500 time steps. This does not mean that PPWS with push probability = 0.5 has better performance than flooding, it actually indicates that the change in the performance of PPWS with probability=0.5 at 1500 time steps is higher than flooding when one unit of simulation time is increased.

For the overheads, we can observe that different push probability values result in different costs. This is because different push probabilities influence the chance of nodes to push information to other nodes. The higher the probability is, the higher the chance of nodes pushing information to other nodes which also incurs high overhead costs. Figure 6.32 shows the overhead cost profile which has the same pattern as the information profile in Figure 6.27.

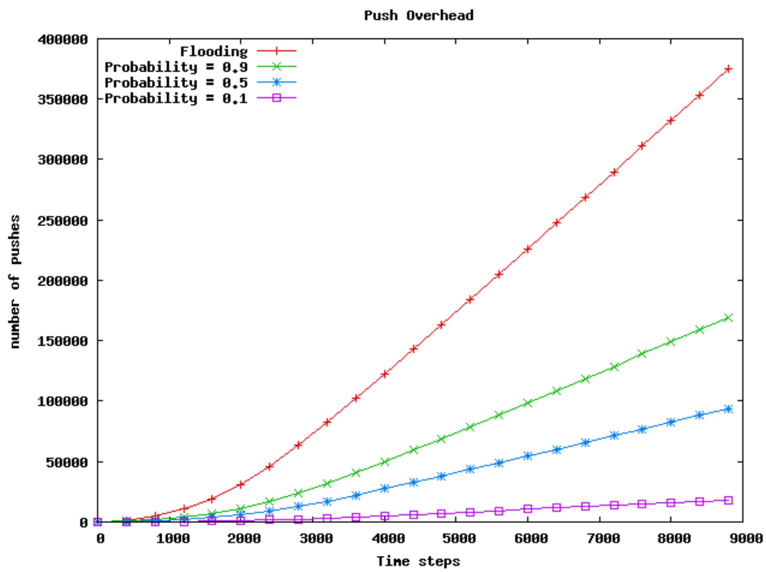


Figure 6.32. Average Push Overhead Costs for Push Probability with Structure using Random Waypoint

Looking at Figure 6.33, the marginal overhead costs increase very quickly when a unit of simulation time is increased at the early stage. This is because more nodes are involved in pushing information as the number of nodes with a artifact accelerate when one unit of time is increased. The big gap of marginal overhead costs between flooding and PPWS with 0.9 is because PPWS uses social structure to guide the push process which reduces

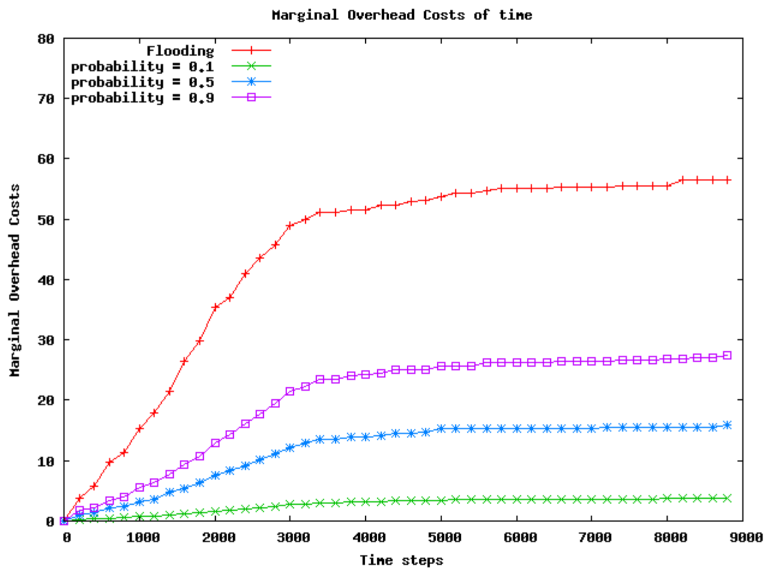


Figure 6.33. Marginal Overhead Costs of time for PPWS using Random Waypoint

the number of pushes among nodes.

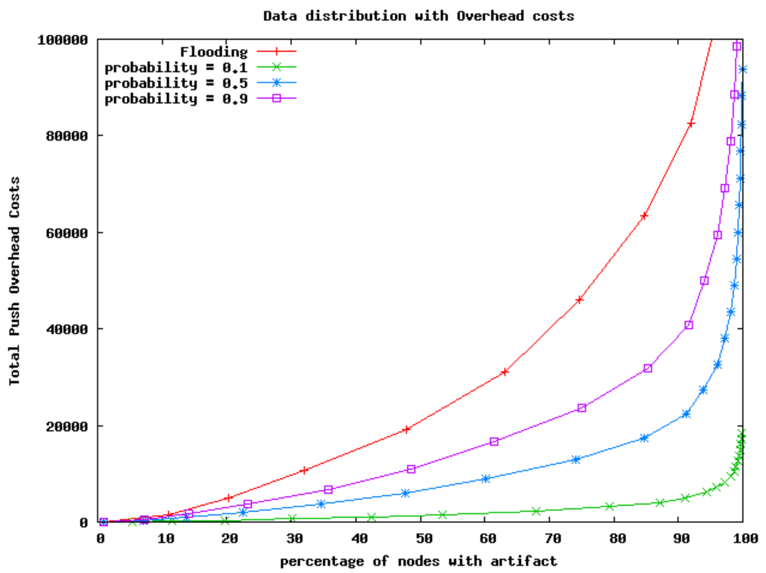


Figure 6.34. Overhead Costs vs Node with artifact for PPWS using Random Waypoint

Based on Figure 6.34, we can observe that the PPWS with probability=0.1 has the lowest overhead costs as compared to the other approaches in the graph. This is because PPWS with probability=0.1 has a small amount of overhead cost. However, because time is not considered in this figure, it does not mean that PPWS with probability=0.1 is the best approach. It has low overhead costs but it has a long delay period in making information available to all nodes in the network (refer Figure 6.30).

6.6.2.3 Push Probability with Structure using Gauss Markov

In Figure 6.35, we observe that there are very small changes in information performance (i.e. number of nodes with artifact) when different push probability is used. This is because the meeting frequency between different nodes is very high in the Gauss Markov mobility model. Therefore, there are more chances for nodes to discover and push information to different nodes even though different push probabilities values are used.

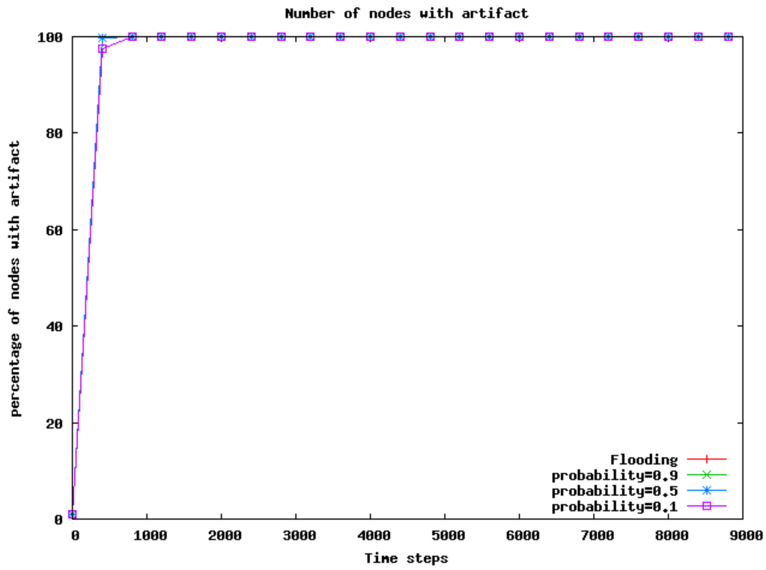


Figure 6.35. Information Profile for Push Probability with Structure using Gauss Markov

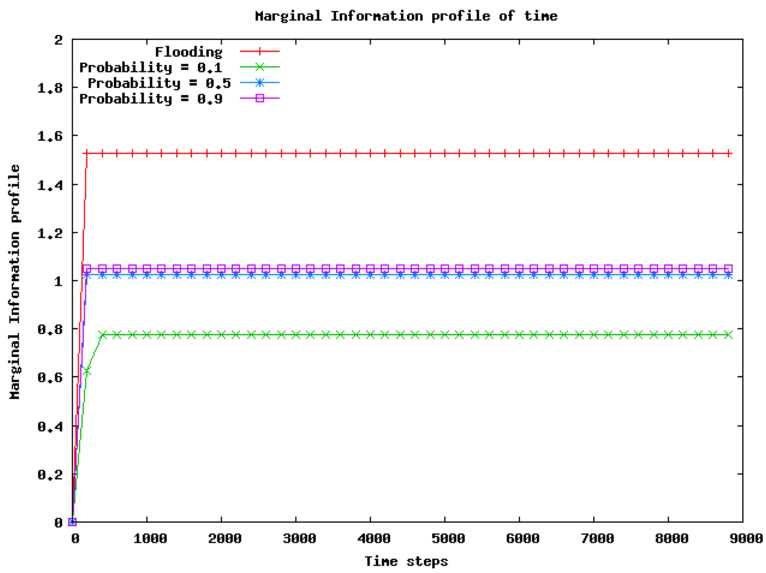


Figure 6.36. Marginal Information profile of time for PPWS using Gauss Markov

Figure 6.36 shows the change in performance (number of nodes with an artifact) when a unit of simulation time is increased. As can be seen from the figure, the performance accelerates very quickly at the very early stage. The higher the push probability is, the faster the information is available to all nodes. This is because with a high probability the nodes have a high opportunity to forward information. The social structure does not confine the nodes to forwarding information because nodes have a big opportunity to meet many different nodes. Therefore it is easy for nodes to form a social structure which consists of different nodes.

In contrast, even though the different push probability values have a small effect on the information profile there is a significant gap in overhead costs. This is because the probability value determines the ability of nodes to push information. Thus, the PPWS with high probability has high overhead costs compared to the PPWS with low probability. The overhead cost is presented in Figure 6.37.

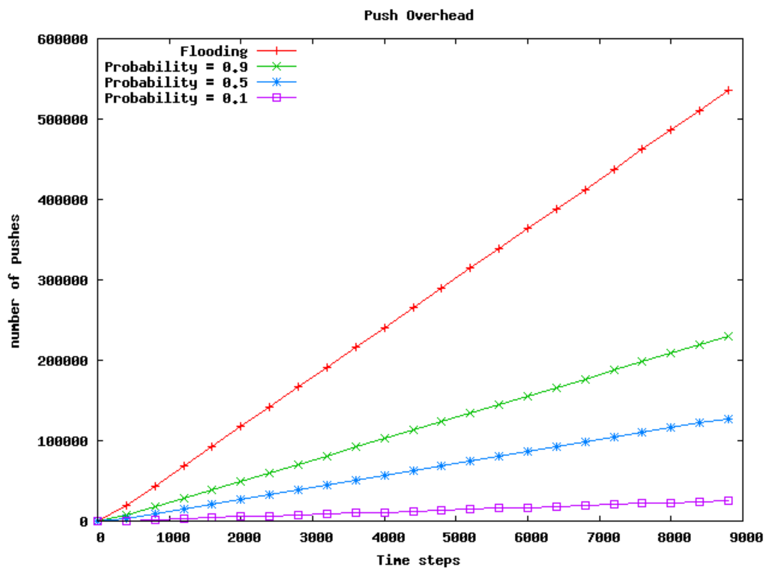


Figure 6.37. Average Push Overhead Costs for Push Probability with Structure using Gauss Markov

Based on Figure 6.38, the overhead costs increase very quickly at the early stage when one unit of simulation time is increased. This is because more nodes are involved in forwarding information as the number of nodes with artifacts increases drastically at that period of time. The big gap between flooding and the PPWS with probability=0.9 in marginal overhead costs is because in flooding nodes greedily pushes information at every

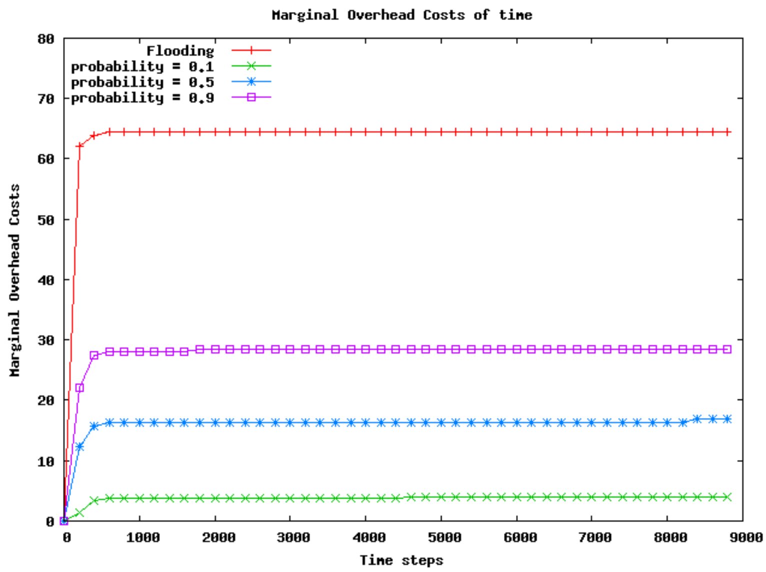


Figure 6.38. Marginal Overhead Costs of time for PPWS using Gauss Markov

opportunity whereas in PPWS the pushing process is controlled by the social structure and the push probability.

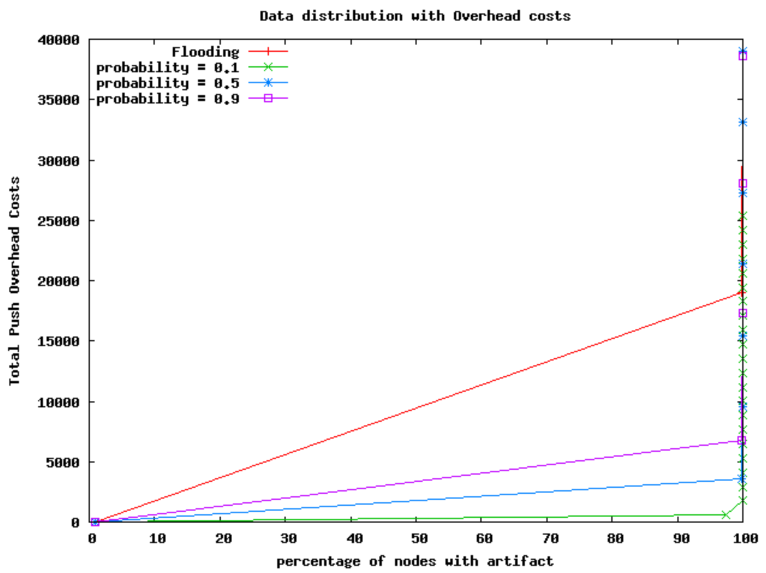


Figure 6.39. Overhead Costs vs Node with artifact for PPWS using Gauss Markov

Based on Figure 6.39, we can observe that when the overhead costs increase the percentage of nodes with an artifact is also increased. Using the Gauss Markov mobility model a smaller amount of overhead is incurred in disseminating the artifact to all nodes in comparison with other mobility models (i.e. Random Walk and Random Waypoint). For example looking at the flooding approach, Gauss Markov results in approximately

20,000 overhead costs to reach 100% performance where Random Walk results in more than 100,000 overhead costs (refer figure 6.34). This is because the Gauss Markov mobility model offers a better chance for nodes to meet different nodes at the early stage of simulation. Therefore, a small amount of resources is enough to spread information very quickly to all nodes.

6.6.2.4 Push Probability with Structure using D-GM

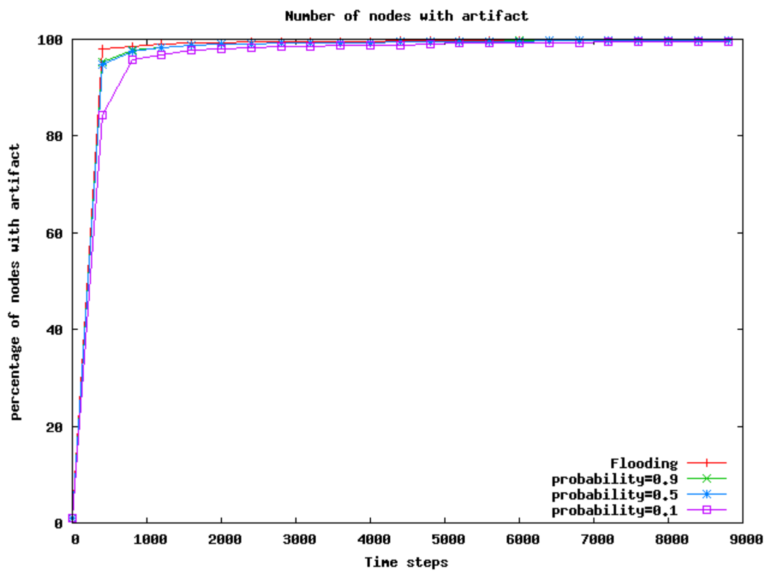


Figure 6.40. Information Profile for Push Probability with Structure using D-GM

As in PPWS with D-GM, increasing the push probability has a small impact on the information profile. This is because D-GM mobility model offers a better chance for nodes to meet different nodes at an early stage. However when the performance is above 80% (Figure 6.40), the performance of PPWS with probability=0.1 is decreased. This is because a node has a small chance to push information when it is in range with other nodes that have not yet possessed information. In comparison to PPWS with high probability, the chance of pushing information at every meeting is high.

Looking at Figure 6.41, we can see that the performance accelerates very quickly when one unit of time is increased. This shows that more nodes discover information at the early stage. This is due to the fact that the D-GM mobility model creates more chances for the nodes to discover different nodes at different time steps. This helps to spread the information quickly to all nodes in the network. The different level in marginal information

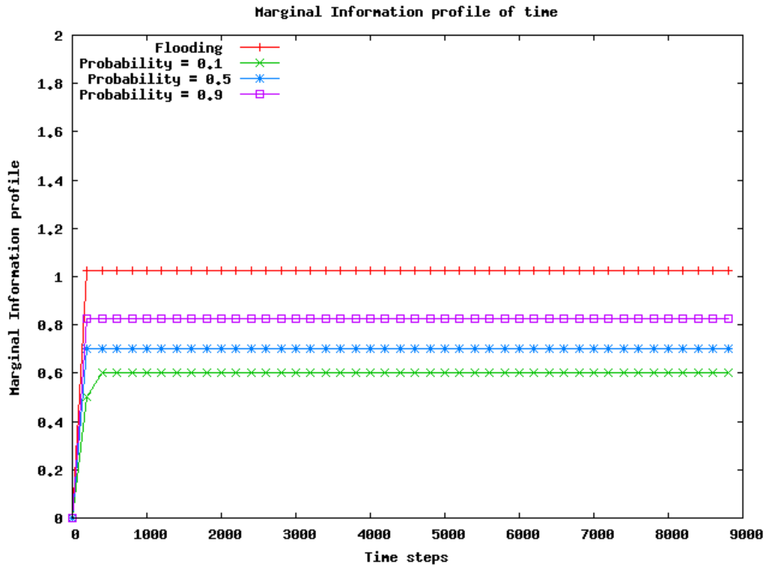


Figure 6.41. Marginal Information profile of time for PPWS using Gauss Markov

profile is shown in Figure 6.41 because of the push probability behavior.

In Figure 6.42, the overhead costs increase over time. This is because more nodes are involved in pushing information as more nodes possess information over time. Even though the different push probability values have a small effect on the information profile performance, it has significant impact on overhead costs. This is because with a high push probability, more resources are needed as more nodes are involved in pushing information over the time.

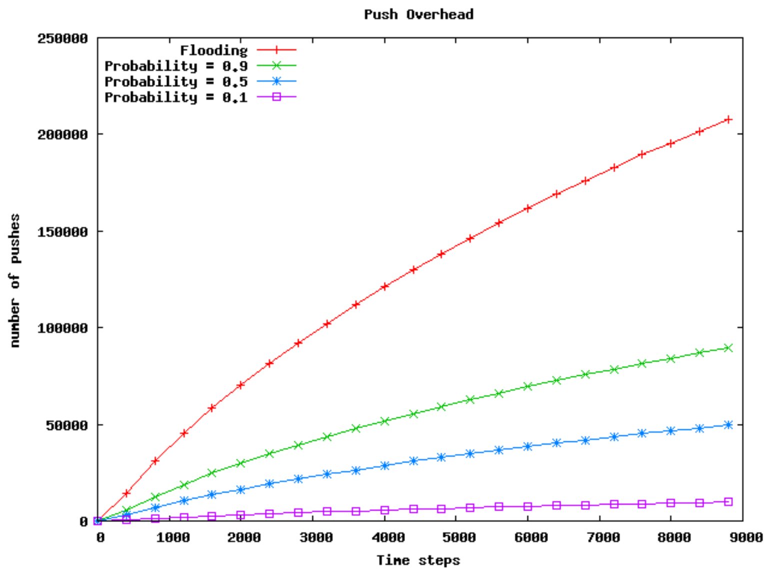


Figure 6.42. Average Push Overhead Costs for Push Probability with Structure using D-GM

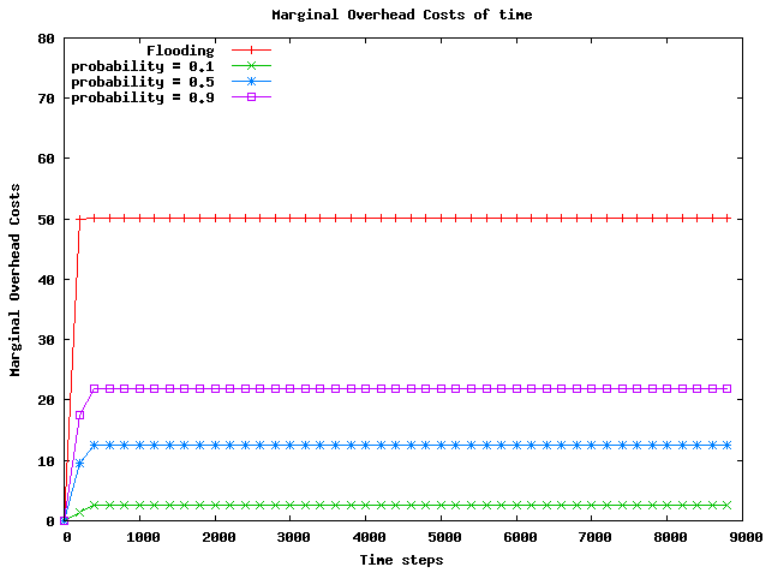


Figure 6.43. Marginal Overhead Costs of time for PPWS using D-GM

In Figure 6.43, the overhead costs accelerate very quickly at the early stage. This shows that nodes are involved in forwarding information as the number of nodes with an artifact increases drastically at that period of time. The gap between flooding and the PPWS with push probability=0.9 is because in flooding nodes are not restricted by any rules to push information. However in PPWS, nodes are confined by the social structure and the push probability. Thus, less resources is required in PPWS when the probability is low.

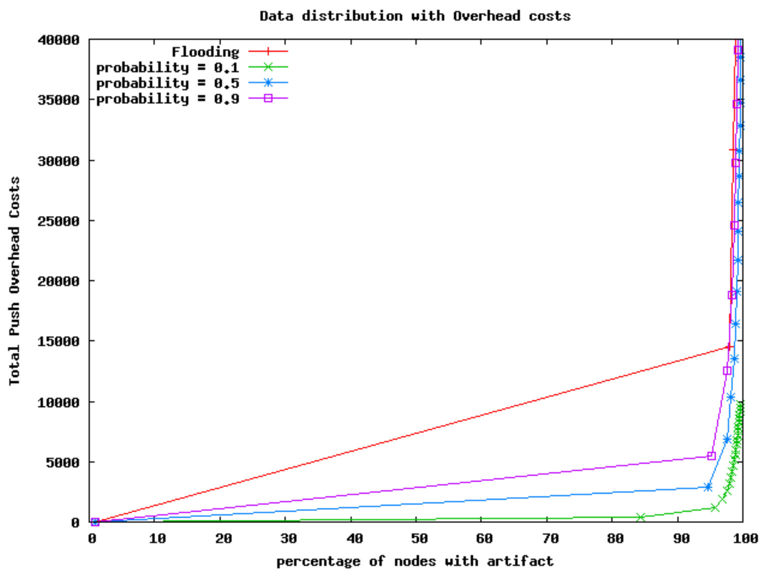


Figure 6.44. Overhead Costs vs Node with artifact for PPWS using D-GM

As can be seen in Figure 6.44, the push probability of 0.1 performs better that others. However because of the time that an artifact is available to all nodes are not considered here, we cannot state that the push probability=0.1 with a lower overhead costs is the best approach.

6.6.3 Results-Push probability without structure

6.6.3.1 Push probability without structure (PPWOS) using Random Walk

The performance of information spreading using PPWOS is determined by different push probability values. As we can see in Figure 6.45, a high push probability value has similar performance as flooding. This is because the nodes with a high possibility almost have push capability the same as flooding where nodes can push information at every meeting opportunity. Therefore, more nodes discover information very early. However, limiting the push frequency slows down the performance of information spreading because nodes have fewer chances to push information at a meeting opportunity.

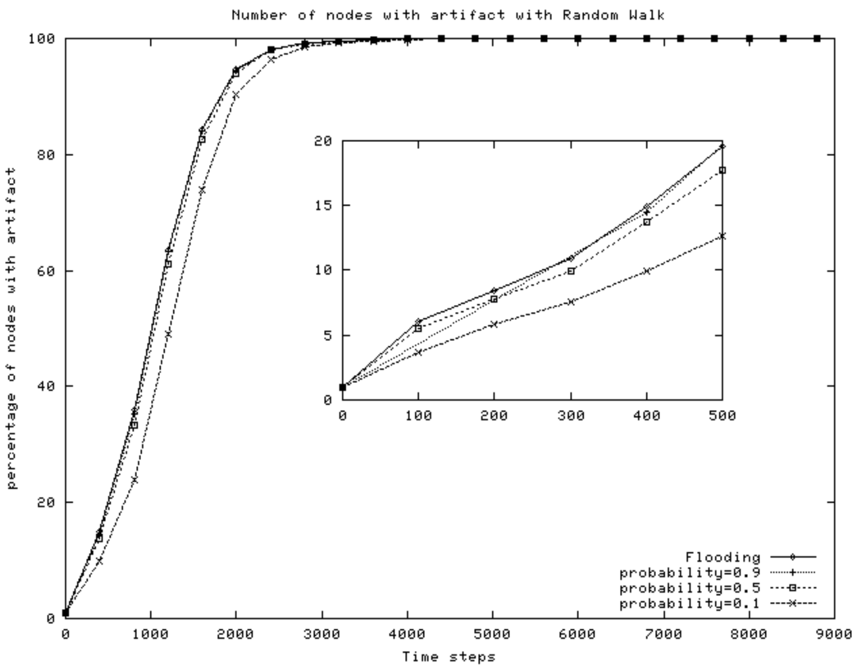


Figure 6.45. Information Profile for Push Probability without Structure using Random Walk

Looking at the marginal information profile of time (Figure 6.46), we can observe that there are different rates of acceleration for PPWOS approach at the early stage.

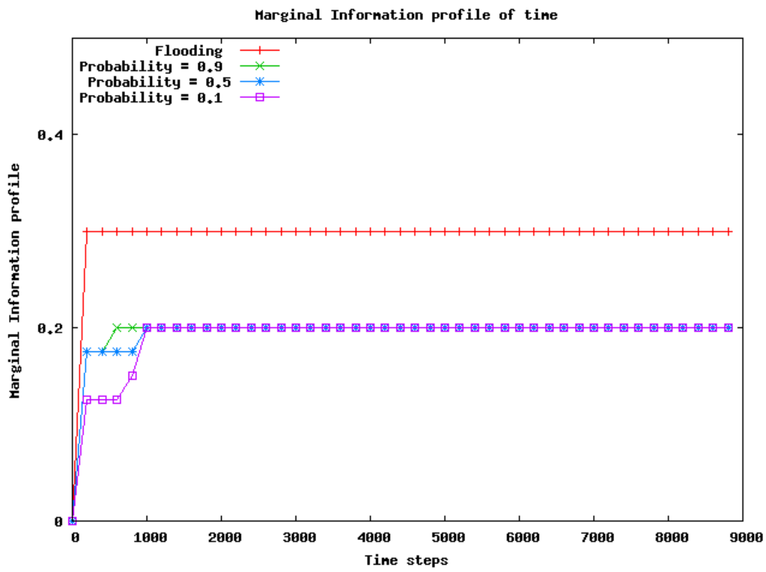


Figure 6.46. Marginal Information profile of time for PPWOS using Random Walk

The PPWOS with high probability (probability =0.9) provides a greater of number of nodes with an artifact when one unit of time is increased as compared to other PPWOS approaches. This is because different push probabilities determine the chances of nodes forwarding information. PPWOS with a low probability has a static marginal information profile at the beginning of the simulation (i.e before 1000 time steps) because at that period of time not many nodes are involved in pushing information.

Figure 6.47 shows the total overhead costs that used by the nodes over time. The usage of resources which is counted as overhead costs are influenced by the number of nodes that are involved in information pushing. As we can observe from the figure, all approaches have lower total overhead costs before 2000 time steps. This is because at that period of time, not many nodes discover information, which prevents them from forwarding information. The overhead costs increase steadily after 2000 time steps. This is because at this point many nodes have discovered information and have started to be involved in pushing information. The gap between the approaches in Figure 6.47 is because of the differences in the ability (chance) to push information.

Based on Figure 6.48, the marginal overhead costs accelerate when one unit of time is increased at the early stage (i.e. before 2000 time steps). This shows the number of nodes discovering information and that are involved in forwarding information effectively at the

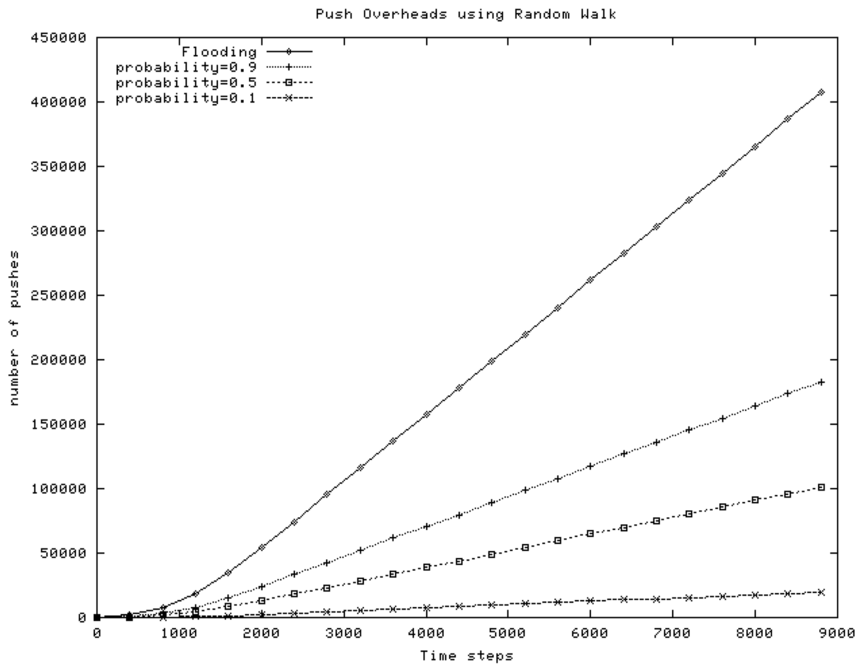


Figure 6.47. Average Push Overhead Costs for Push Probability without Structure using Random Walk

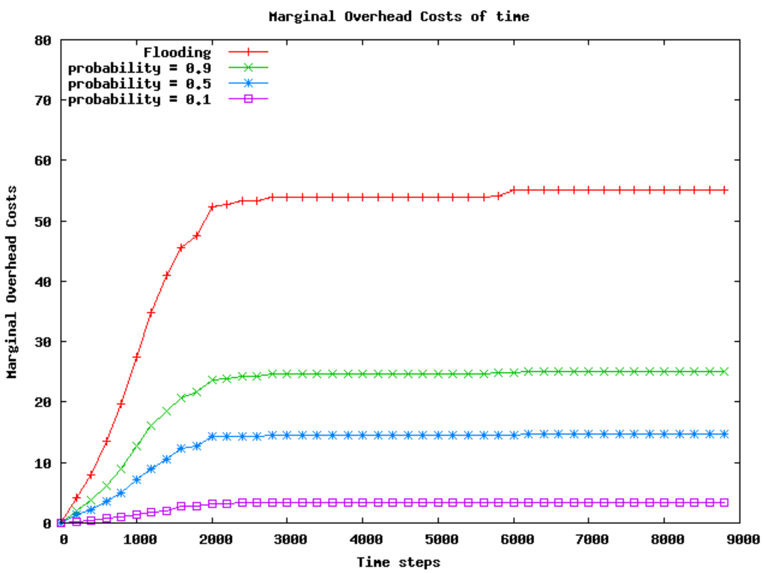


Figure 6.48. Marginal Overhead Costs of time for PPWOS using Random Walk

early stage. The rate of change in cost depends on the ability of nodes to push information to each other. For example, the PPWOS with low push probability (probability = 0.1) has the lowest overhead costs in Figure 6.48 and it has very limited possibilities to push information.

Figure 6.49 shows the relationship between overhead costs and the percentage of nodes

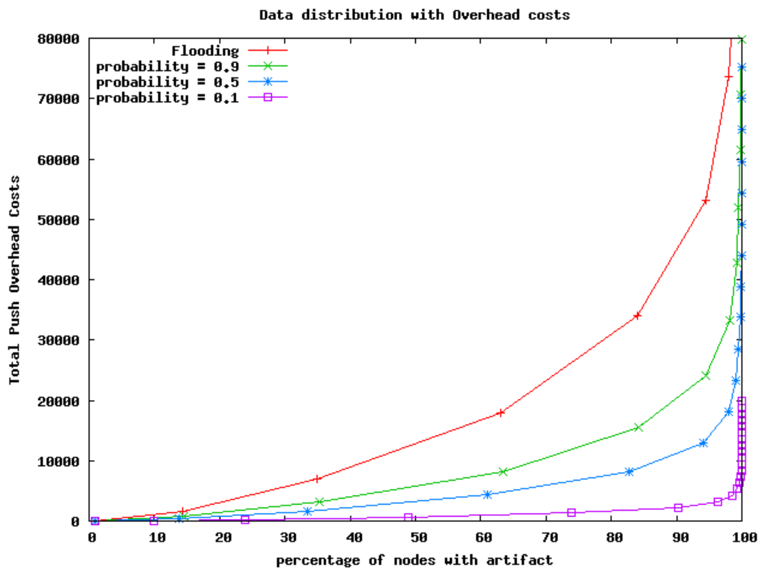


Figure 6.49. Overhead Costs vs Node with artifact for PPWS using Random Walk

with an artifact. Note that the graph is plotted without considering how quickly the information is available to the nodes. The figure investigates the effect of the overhead costs to the percentage of nodes with an artifact. As can be seen, with small overhead costs, it is possible to disseminate information to all nodes in the network. But it requires an amount of delay to accomplish that task. Flooding disseminates information very quickly but suffers from a massive overhead cost. From the figure we can see that the flooding approximately required to push 30,000 times to have 80% nodes with an artifact and PPWOS with push probability=0.9 only needs about 15,000 times pushing information. This explains that there is a possibility of reducing the cost in flooding while maintaining the performance because the performance of PPWOS with probability=0.9 is very close to flooding performance (see Figure 6.45).

6.6.3.2 Push probability without structure using Random Waypoint

In Figures 6.50 and Figure 6.45 show the similar effect on the information dissemination performance when different push probabilities are applied. Looking at Figure 6.50, the gap between flooding and PPWOS with push probability=0.9 can be compared as in Figure 6.45. In the Random Waypoint mobility model nodes have to stop at a certain location for a period of time before they can move to the next destination. This affects the meeting frequency between nodes. Therefore, the chance of pushing information to different nodes also affected. As a result, the information distribution using Random Waypoint is slightly slow as compared to Random Walk mobility model.

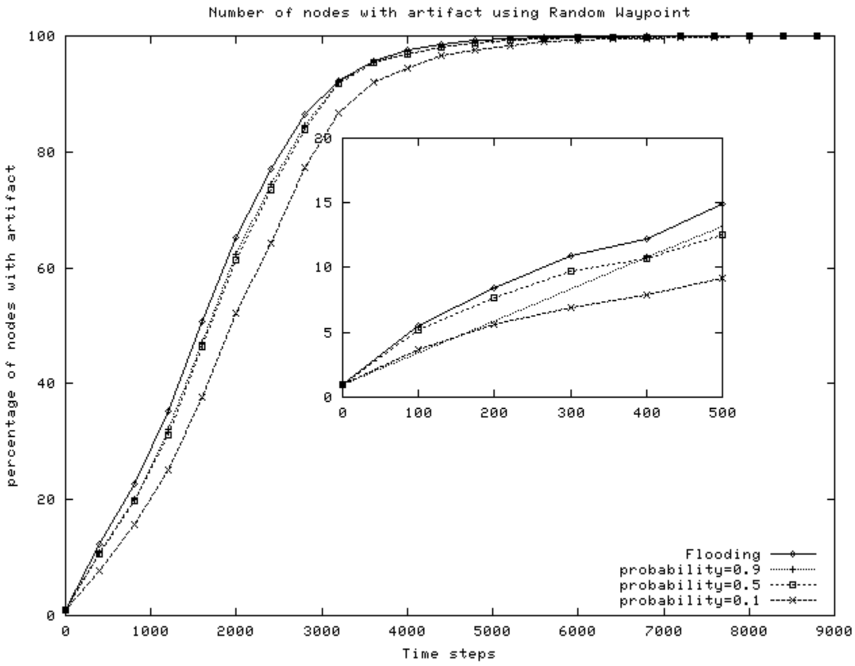


Figure 6.50. Information Profile for Push Probability without Structure using Random Waypoint

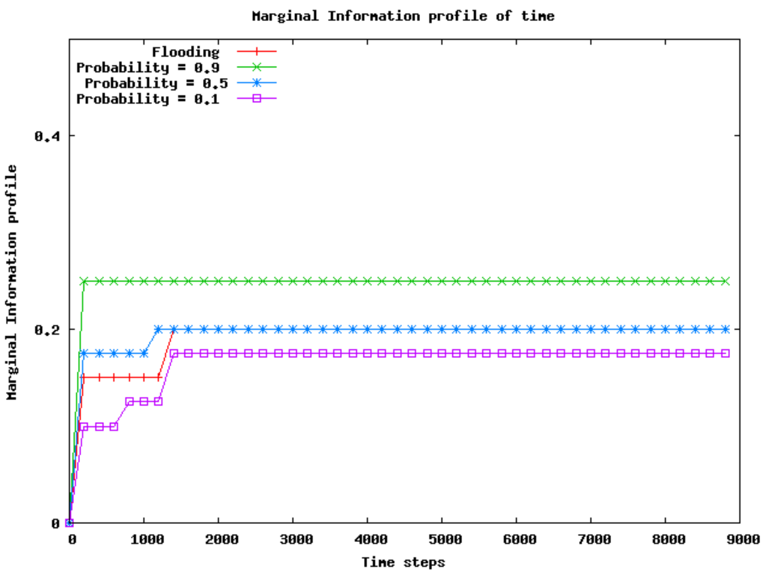


Figure 6.51. Marginal Information profile of time for PPWOS using Random Walk

Figure 6.51 shows the change on performance (i.e. the percentage of nodes with an artifact) when one unit of time is increased. As can be seen from the figure, the PP-WOS with probability=0.9 and probability=0.5 have marginal information profiles that are higher than flooding at the early stage. This is does not mean that PPWOS with probability=0.9 and probability=0.5 are better than flooding, but this indicates that the there is a change in the performance for PPWOS approach at the time which is better than flooding. This is because flooding has constant marginal information profile.

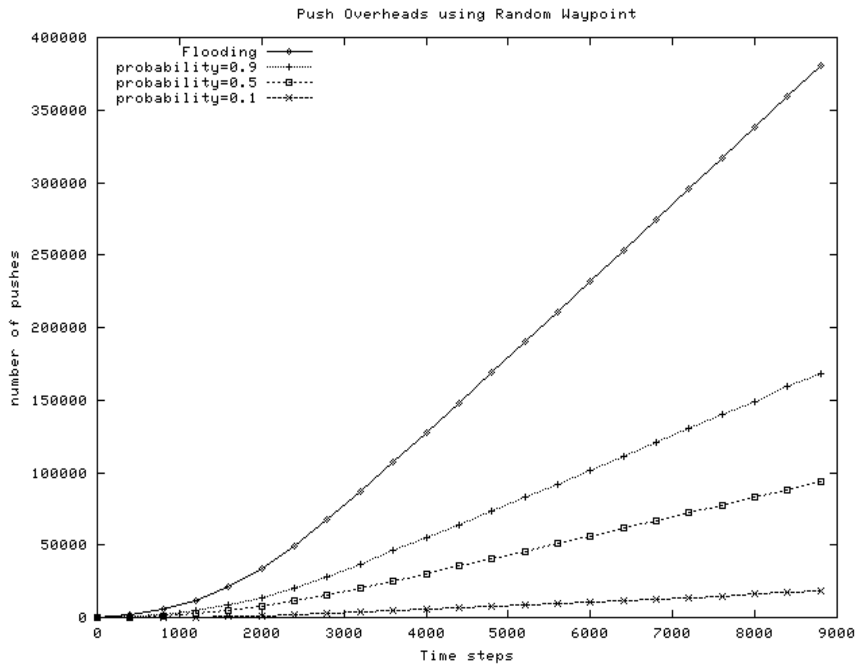


Figure 6.52. Average Push Overhead Costs for Push Probability without Structure using Random Waypoint

Figure 6.52 shows the overhead costs for different PPWOS using the Random Waypoint mobility model. Here the total overhead costs increases as simulation time is increased. This is because the number of nodes with an artifact increases over time. Therefore, more nodes are involved in forwarding information which causes the total overhead costs to also increase over time. The total overhead costs gap among the PPWOS approaches indicates that each of the approaches has different frequency of pushing. The high probability has the most total overhead costs as the probability of pushing information is very high.

Looking at the 6.53, the marginal overhead cost increases quickly when a unit of simulation time is increased. This is because the number of nodes involved in pushing information increases over time. The gap between the approaches is because each approach has its own predefined pushing frequency which determines the capability of nodes to push at every unit of time. Therefore, as expected the PPWOS with high probability will have

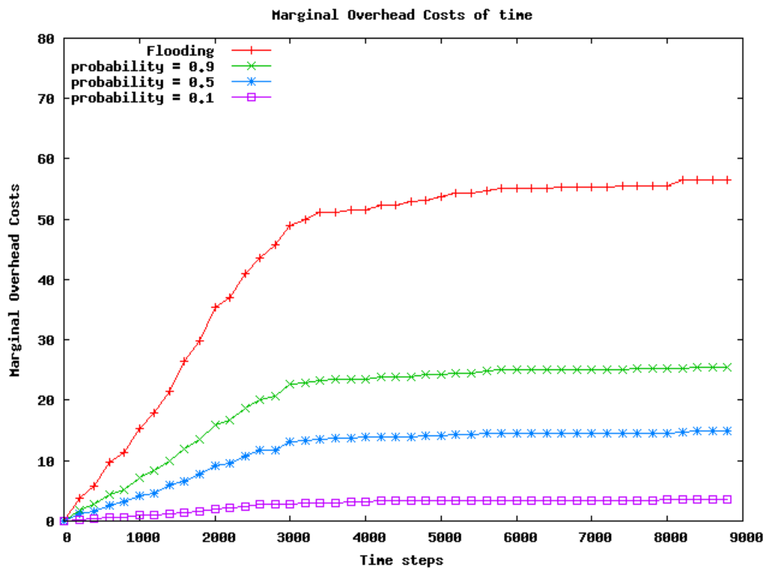


Figure 6.53. Marginal Overhead Costs of time for PPWOS using Random Waypoint

a high marginal cost at the early stage as information is available more quickly to the nodes.

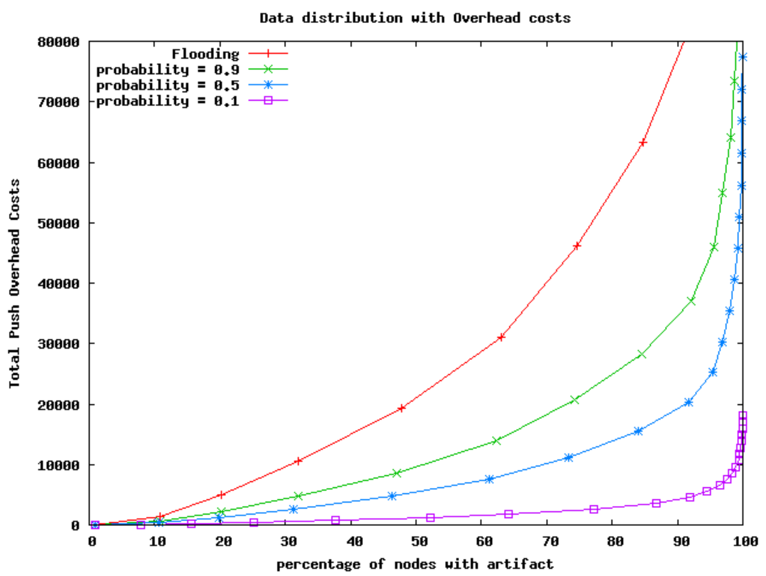


Figure 6.54. Overhead Costs vs Node with artifact for PPWOS using Random Waypoint

Figure 6.54 shows the relationship between overhead costs and the percentage of nodes with artifact. Note that the graph is plotted without considering how quickly the information is available to the nodes. This figure is mainly to investigate what is the effect of the overhead costs to the percentage of nodes with artifact using Random Waypoint. As can be seen from the figure, it is possible to disseminate information to all nodes in the network with small overhead costs. But it requires amount of delay to accomplish that task as with the PPWOS with probability=0.1. Flooding disseminates information very quickly however it suffers from very high overhead costs. From the figure we can see that flooding uses approximately 80,000 in costs to reach 80% of nodes with an artifact but PPWOS with probability=0.9 only needs about 20,000 in costs. To have a high data dissemination performance, cost is always a constrain however it can be reduced by controlling the push mechanism.

6.6.3.3 Push probability without structure using Guass Markov

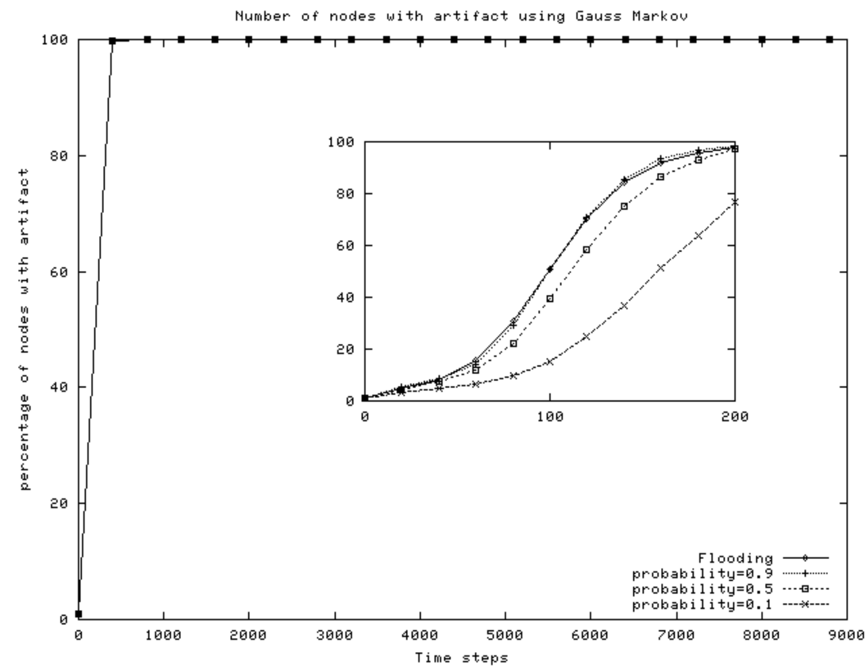


Figure 6.55. Information Profile for Push Probability without Structure using Guass Markov

As can be seen from Figure 6.55 varying the push probability value has small effect on information spreading performance when using the Gauss Markov model. This is due to the fact that the Guass Markov mobility model creates more chances for nodes to meet different nodes at the early stage. Therefore, the chance of information being available very quickly to all nodes is very high even though the push probability is very small (i.e. probability=0.1). Looking at the small scale graph in 6.55, using a small push probability still has impact on the information dissemination performance. This is because different push probabilities determine the push frequency. So, with the high push probability more nodes discover information at the early stage (i.e. within 200 time steps).

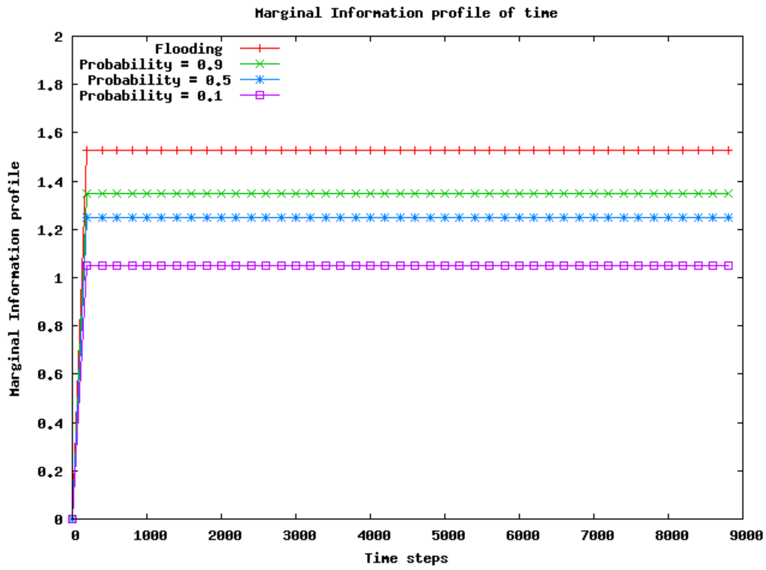


Figure 6.56. Marginal Information profile of time for PPWOS using Gauss Markov

Figure 6.56 shows the change in performance (i.e. the percentage of nodes with artifact) when one unit of time is increased. As can be seen from the figure, the marginal performance accelerates very quickly at the early stage and there is little to distinguish different push probability. The gap between approach in the Figure 6.56 shows the impact of different setting of push probability.

In contrast, even though the different push probability values have a small effect on the information profile there is a significant gap in overhead costs. This is because nodes with a high probability have a big chance to push more frequently which results in more costs. Thus, this explains how the PPWOS with high probability has high overhead costs compared to the PPWOS with low probability as shown in Figure 6.57.

Based on Figure 6.58, the overhead costs increase very quickly at the early stage when one unit of time is increased. The figure also shows that in all cases a constant increase is achieved beyond the very early stage. The big gap between flooding and the PPWOS with probability=0.9 is because in flooding nodes greedily push information at ever opportunity whereas in PPWOS the information pushing is confined by the push probability.

Based on Figure 6.59, we can observe that the percentage of nodes with an artifact increases quickly even though a small amount of overhead cost is used. This is because the

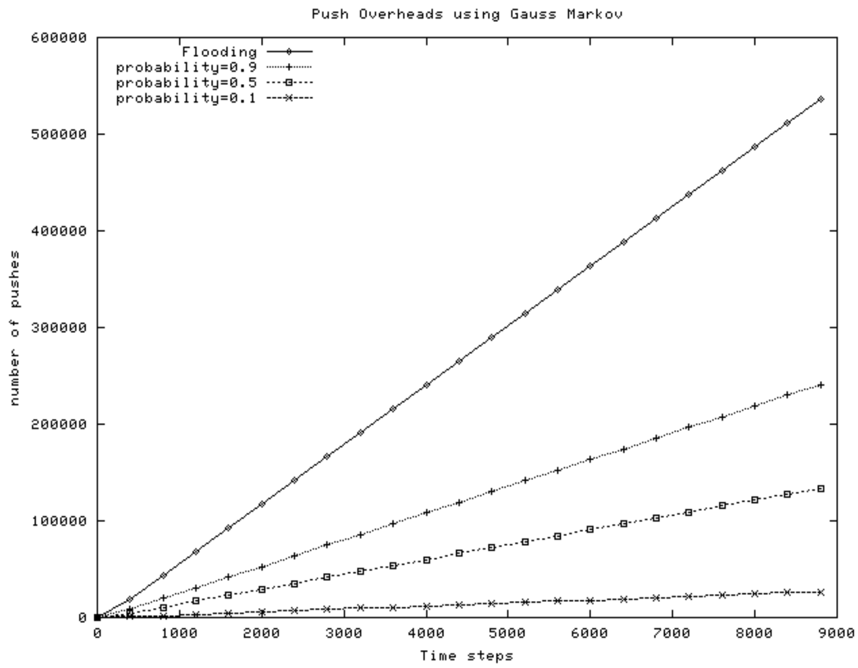


Figure 6.57. Average Push Overhead Costs for Push Probability without Structure using Gauss Markov

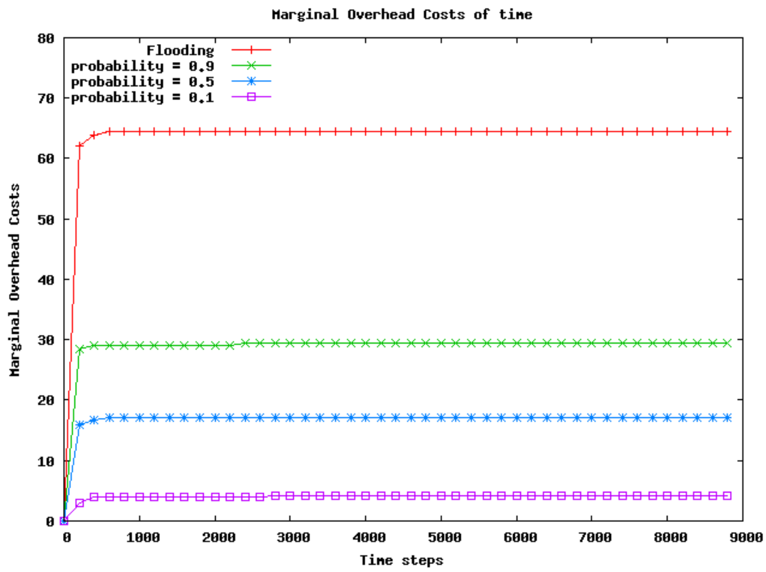


Figure 6.58. Marginal Overhead Costs of time for PPWOS using Gauss Markov

Gauss Markov mobility model creates more opportunity for nodes to discover information at the early stage. For example, looking at the flooding approach, Guass Markov uses approximately 20,000 units of overhead costs to achieve 100% performance (i.e percentage of nodes with an artifact) whereas Random Walk needs more than 80,000 units of overhead costs (Figure 6.49) to achieve the same information dissemination.

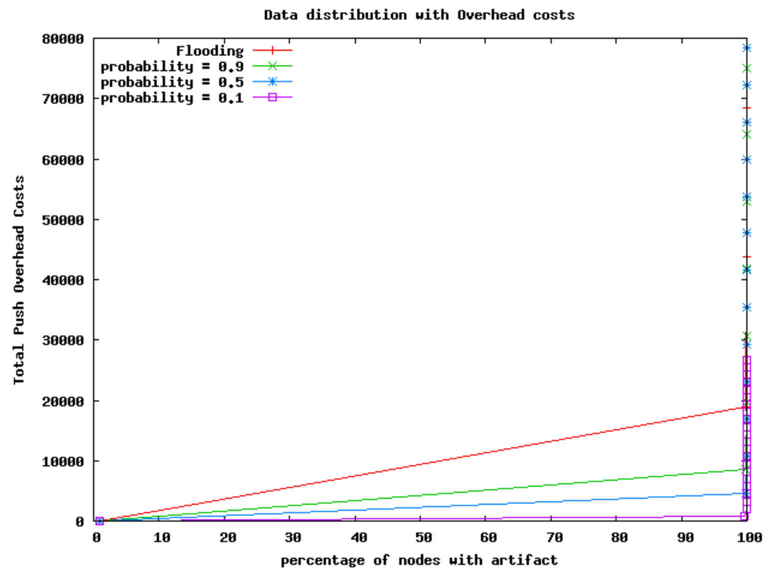


Figure 6.59. Overhead Costs vs Node with artifact for PPWOS using Gauss Markov

6.6.3.4 Push probability without structure using D-GM

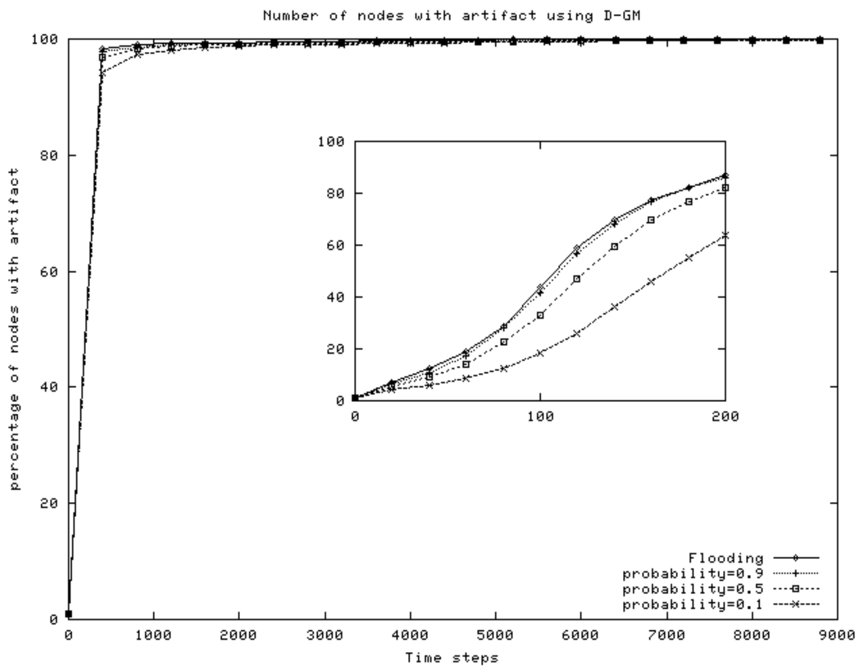


Figure 6.60. Information Profile for Push Probability without Structure using D-GM

In Figure 6.60, a high value of push probability brings the information spreading performance close to flooding performance. Because of more opportunity for nodes in seeing each other frequently using D-GM mobility model, setting the Push Probability further close to zero has a small impact on information spreading performance at the early stage.

This is can be seen in the small scale graph in Figure 6.60. The gap between PPWOS approaches that shows in the small scale graph is because the nodes are confined by the push probability value which results in the PPWOS with low push probability having slightly more delay in making the information available to all nodes.

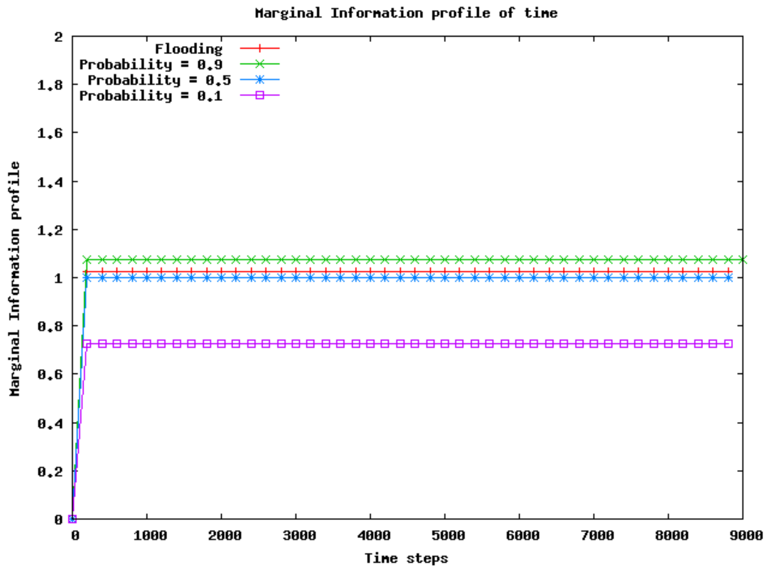


Figure 6.61. Marginal Information profile of time for PPWOS using DGM

Looking at Figure 6.61, we can see that the performance accelerates very quickly when a unit simulation time is increased. This shows that more nodes discover information at the early stage. This is due to the fact that D-GM mobility model creates more chance for nodes to discover information by meeting different nodes. This helps the information spreads quickly to all nodes in the network. The different level in marginal information profile between different PPWOS profile is because of the different push probability values. In Figure 6.61, the marginal performance of PPWOS with probability = 0.9 is slightly higher than flooding. This shows that PPWOS approach has greater number of nodes that discover an artifact at the early stage when a unit of time is increased as compared to the Flooding approach .

Figure 6.62 shows the overhead costs for different PPWOS using D-GM mobility model. As can be seen from the figure, the total overhead cost is increases as the time is increased. This is because the percentage of nodes with artifacts increase over time. Consequently, more nodes are involved in forwarding information which eventually increasing the total overhead costs over time. The total overhead costs gap among the PPWOS approaches

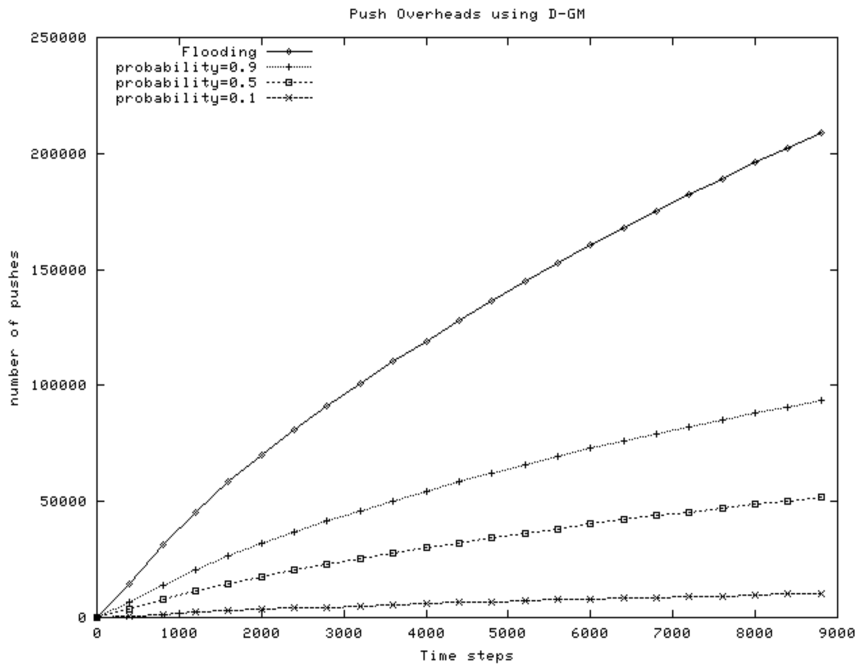


Figure 6.62. Average Push Overhead Costs for Push Probability without Structure using D-GM

is because each of the approaches has a different frequency of pushing. Therefore, as we expected the PPWOS with a high probability has the highest total overhead cost as it has more opportunity to forwards information.

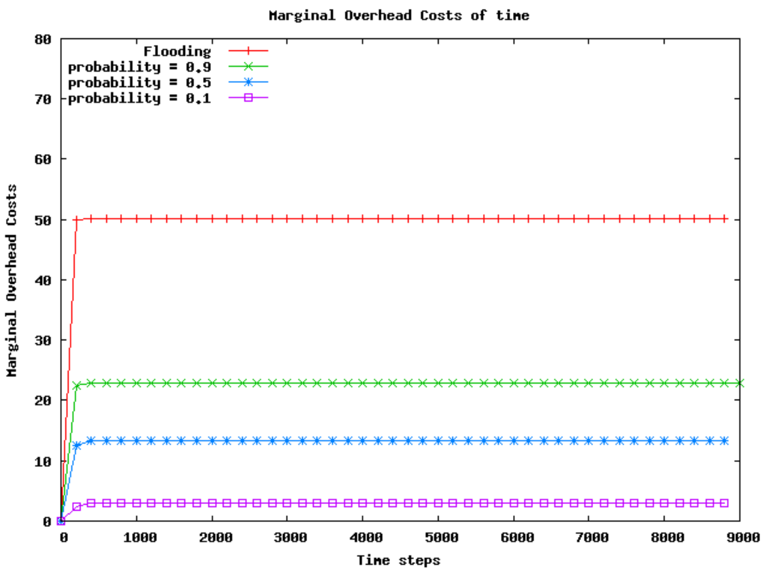


Figure 6.63. Marginal Overhead Costs of time for PPWOS using D-GM

Based on Figure 6.63, the overhead costs increase very quickly at the early stage when one unit of time is increased. This is because more nodes are involved in forwarding information as the number of nodes with artifacts increase drastically at the early stage. The big gap between flooding and the PPWOS with probability=0.9 is because in flooding nodes greedily push information at every opportunity whereas in PPWOS the information pushing is controlled by the push probability setting. Therefore at every time step PPWOS uses less pushing than flooding uses.

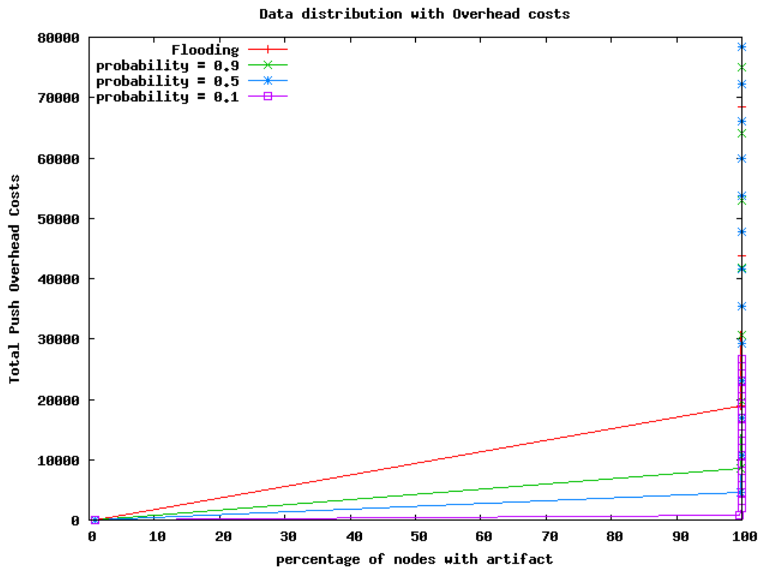


Figure 6.64. Overhead Costs vs Node with artifact for PPWOS using D-GM

Based on Figure 6.64, we can observe that the percentage of nodes with an artifact increases quickly even though a small amount of overhead cost is used. This is because in D-GM mobility model creates more opportunity for nodes to discover information at the early stage. For example looking at the flooding approach, D-GM needs less than 20,000 overhead costs to achieve 100% performance (i.e percentage of nodes with an artifact) whereas Random Walk needs more than 80,000 overhead costs (refer to Figure 6.49). This shows that the number of pushes is directly influence the performance of information dissemination. Overhead cost is a trade-off in order to achieve a better information dissemination.

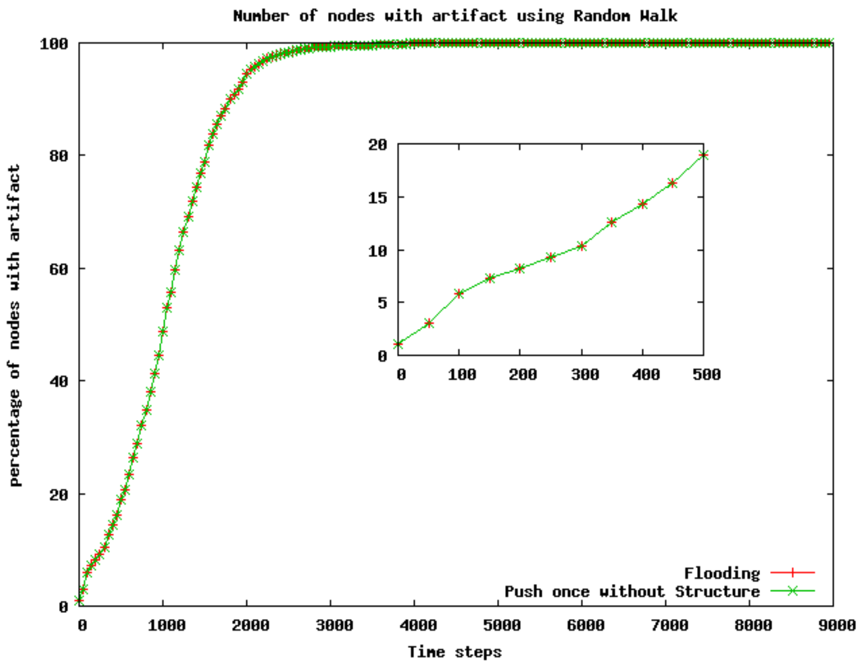


Figure 6.65. Information Profile for Push Once without Structure using Random Walk

6.6.4 Results-Push Once without structure

6.6.4.1 Push Once without structure (POWOS) using Random Walk

In the Push Once approach, a node only pushes information to the nodes that it has not pushed information to before. The advantage of this technique is able to avoid sending information to the same nodes repeatedly. To execute this approach, it is assumed that every node has enough memory space to remember the nodes that have been already pushed. So there is an added requirement for the nodes. Looking at the percentage of node with artifact in Figure 6.65, POWOS has the same performance as flooding. This is because POWOS and flooding discover the same nodes at every simulation steps. POWOS forwards information selectively whereas flooding just forwards information at any opportunities. So, POWOS has similarity with flooding in terms of data dissemination performance but it has better management in reducing the information duplication.

In Figure 6.66, POWOS has the same marginal information profile of time as with flooding. This is because POWOS forwards information the same as flooding in which results in the same percentage of nodes with artifact at every time steps.

Because the POWOS approach only pushes to the nodes that are not in the list,

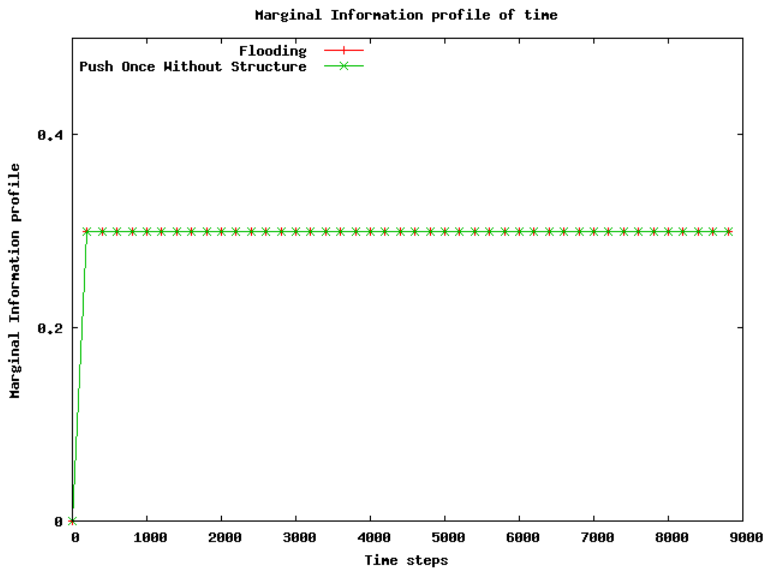


Figure 6.66. Marginal Information profile of time for POWOS using Random Walk

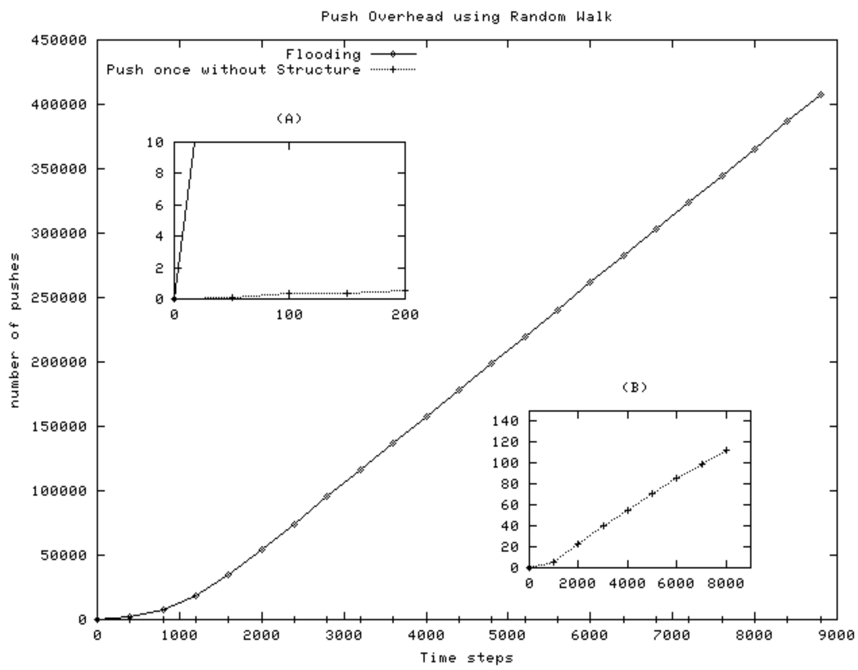


Figure 6.67. Average Push Overhead Costs for Push Once without Structure using Random Walk

therefore the number of pushes involved at every time step reduces over time. This is because the same nodes might be in range within certain periods of time in which flooding is pushing activities still continue even though the nodes already have the information. This causes unnecessary overhead costs in flooding. In the POWOS approach, the unnecessary push in flooding activity is avoided through the interaction history. Therefore, as we expected, the overhead cost of push is far better than flooding. Figure 6.67 shows the total overhead cost for flooding and POWOS.

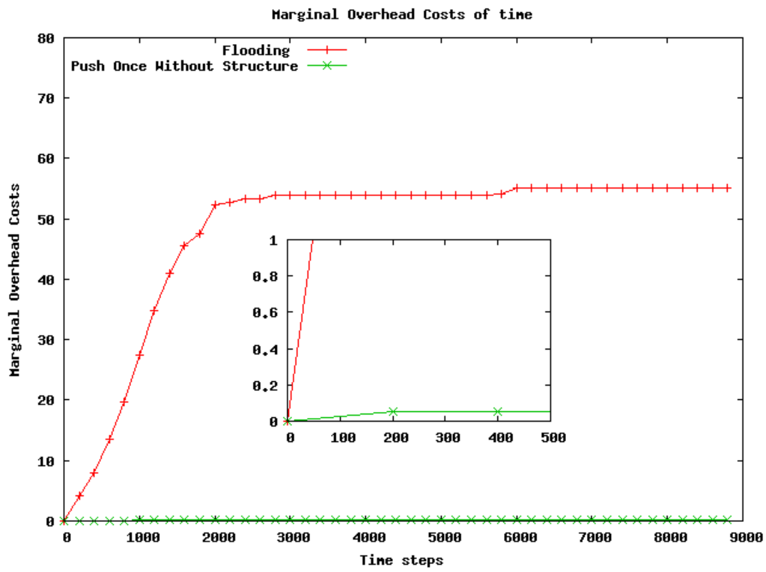


Figure 6.68. Marginal Overhead Costs of time for POWOS using Random Walk

In Figure 6.68 we can observe that there is big gap in marginal overhead costs between POWOS and flooding. This is because in POWOS not many nodes are involved in pushing information as nodes are only pushing information to nodes that it never seen before. Therefore, the change in overhead costs when one unit of time is increased is very small.

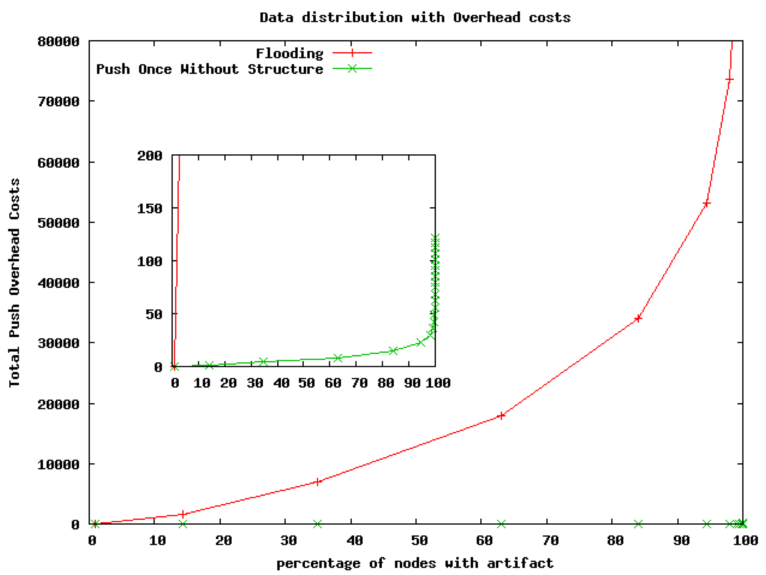


Figure 6.69. Overhead Costs vs Node with artifact for POWOS using Random Walk

Looking at Figure 6.69, we can observe that the there is a big gap in overhead costs and percentage of nodes with artifact between flooding and POWOS. Because both of them have the same performance as shown in Figure 6.65, we can say that POWOS is better that flooding in terms of disseminating information with lower overhead costs.

6.6.4.2 Push Once without structure using Random Waypoint

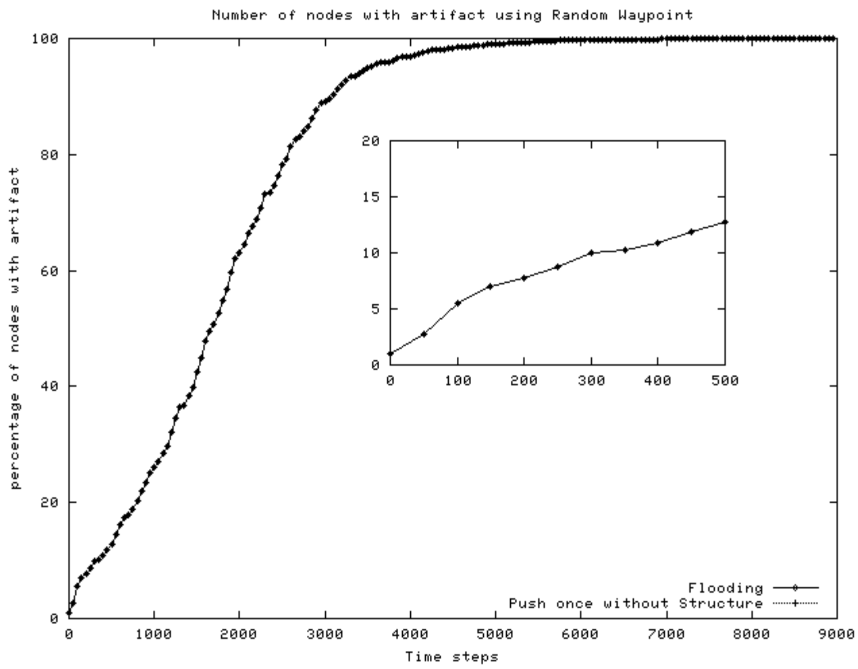


Figure 6.70. Information Profile for Push Once without Structure using Random Waypoint

Figure 6.70 shows that POWOS spreads information similarly to flooding in terms of performance. This is because POWOS has the same forwarding system as flooding. Moreover, POWOS more intelligently pushes information to different nodes that it has never seen before. The information dissemination performance in Figure 6.70 is reduced slightly because Random Walk offers nodes a better interaction frequency compared to the Random Waypoint mobility model.

Figure 6.71 shows the change in percentage of nodes with an artifact when one unit of time is increased. From the figure, obviously we can see that POWOS has the same marginal performance as flooding. This is indicates that POWOS has exactly the same performance as flooding.

In Figure 6.72, we can see that the overhead cost of POWOS is far better than flooding.

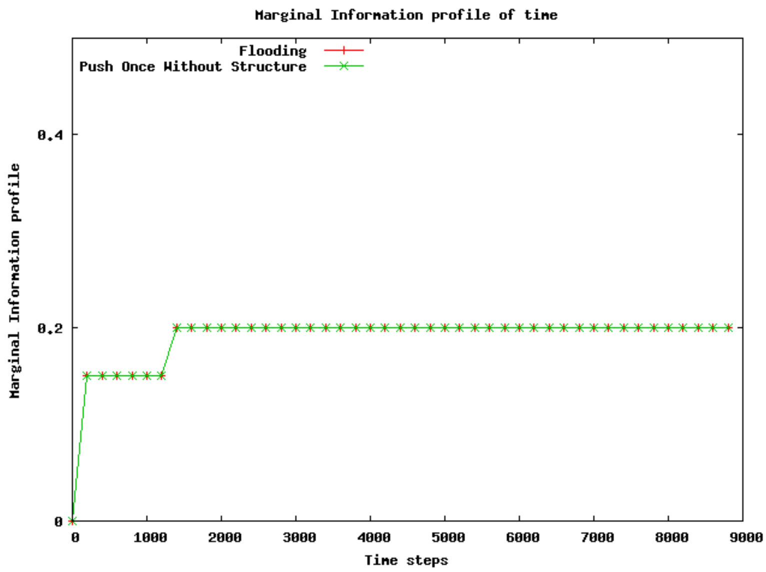


Figure 6.71. Marginal Information profile of time for POWOS using Random Waypoint

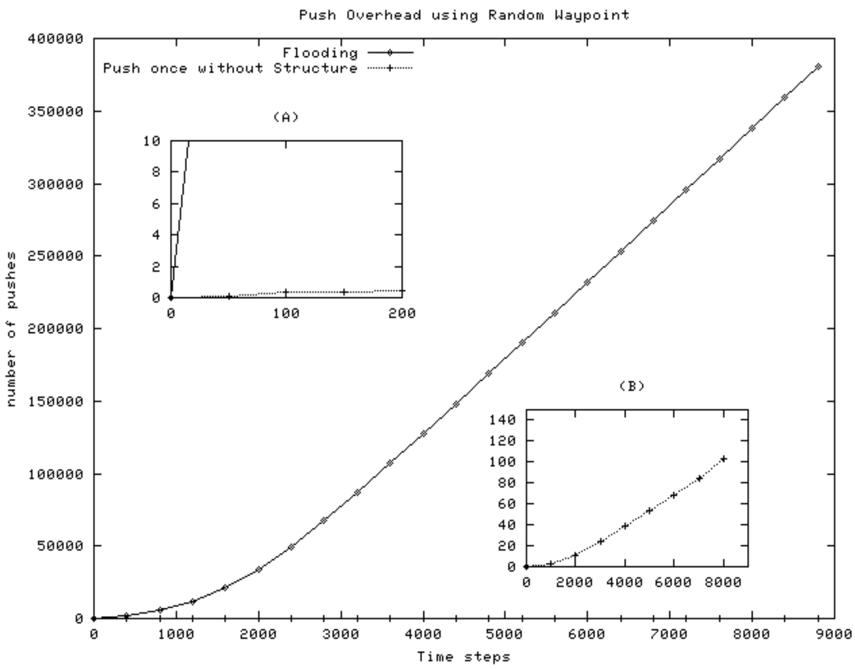


Figure 6.72. Average Push Overhead Costs for Push Once without Structure using Random Waypoint

This is because nodes in POWOS only push information to the nodes that they have never seen before. So, even though the same nodes are in range frequently, a single push is enough to disseminate information. However, in flooding the nodes will push information to each other at every unit of time step.

Based on Figure 6.73, we can observe that there is a big gap in marginal overhead costs

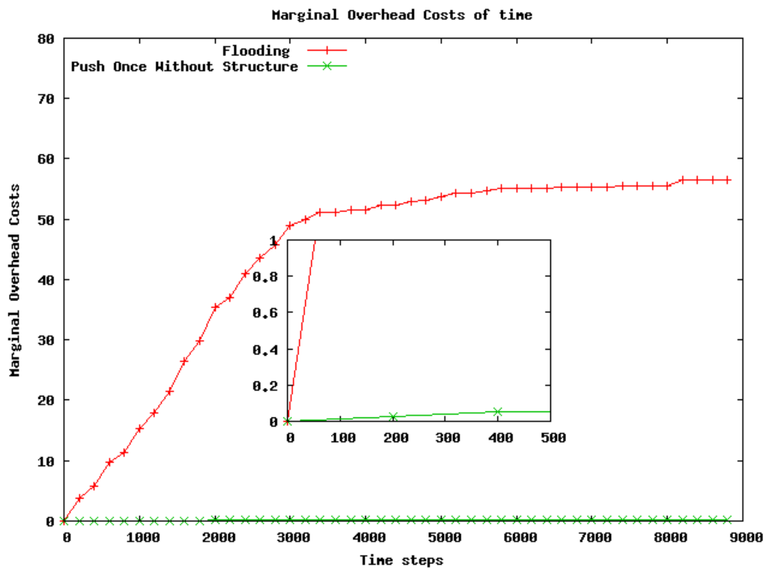


Figure 6.73. Marginal Overhead Costs of time for POWOS using Random Waypoint

between POWOS and flooding. This is because in POWOS not many nodes are involved in pushing information. Therefore, the change in overhead costs when one unit of time is increased is very small. For flooding, the acceleration of marginal cost at the beginning is because many nodes are started to be involved in pushing information.

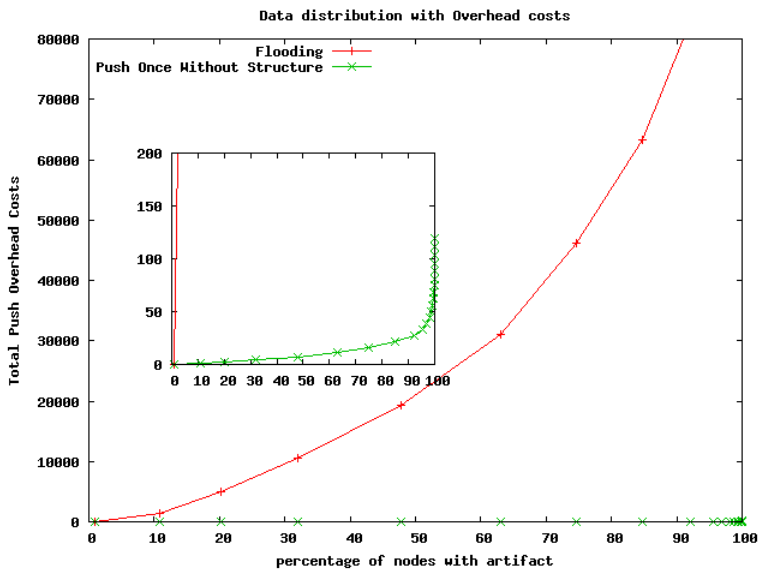


Figure 6.74. Overhead Costs vs Node with artifact for POWOS using Random Waypoint

Looking at Figure 6.74, we can observe that there is a big gap in overhead costs and the percentage of nodes with an artifact between flooding and POWOS. This is because both of the approaches have different ways of pushing information. From the figure, we can say that POWOS is better that flooding because it has a good performance close to flooding and also it has very small overhead costs compared to flooding.

6.6.4.3 Push Once without structure using Guass Markov

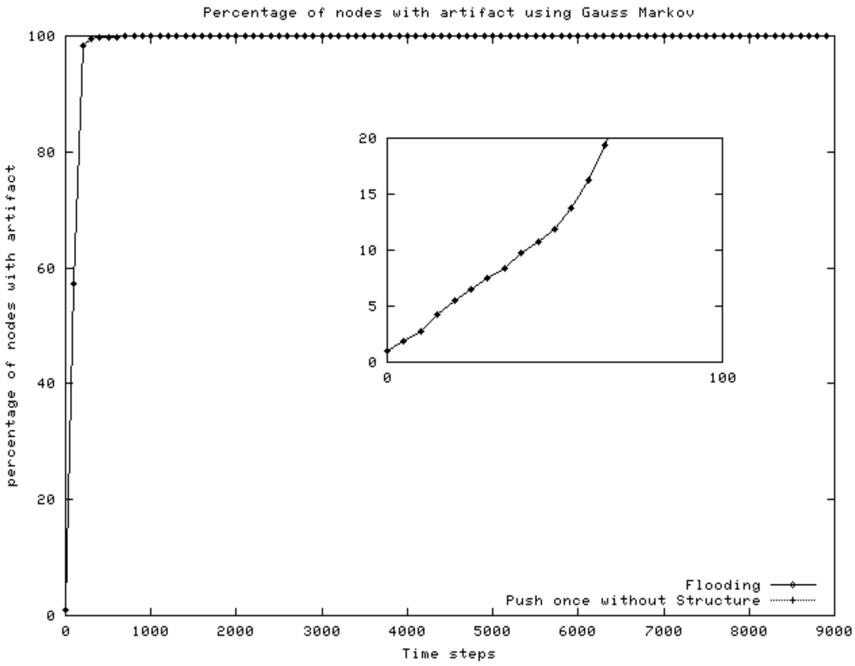


Figure 6.75. Information Profile for Push Once without Structure using Gauss Markov

Using Gauss Markov mobility model, POWOS spreads information to all nodes as quick as flooding. This can be seen from Figure 6.75 where the graph lines are overlapping each other. This is because POWOS forwards information as efficient as flooding where the nodes with an artifact send information to nodes that it has never seen before. So by doing this, POWOS is actually forwarding information similarly to flooding but it is forward information more intelligently.

In Figure 6.76, we see that the chance in percentage of nodes with artifact when one unit of time is increased in both approach (POWOS and flooding). As we expected that POWOS has the same marginal performance as flooding. This is because both of them has the same percentage of nodes with artifact at every time steps.

Because POWOS has system to avoid pushing information to the same nodes, it has small overhead costs in comparison to the flooding approach. This is can be observed from Figure 6.77. The big gap in overhead costs between POWOS and flooding is because

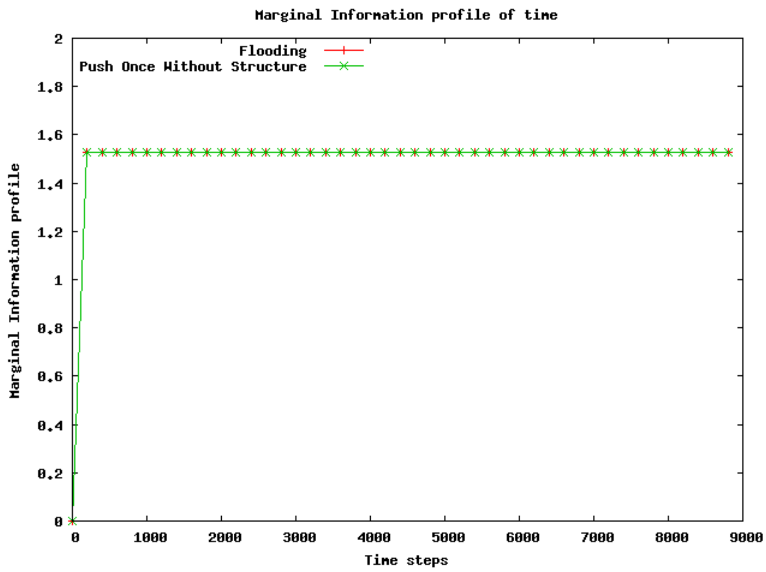


Figure 6.76. Marginal Information profile of time for POWOS using Gauss Markov

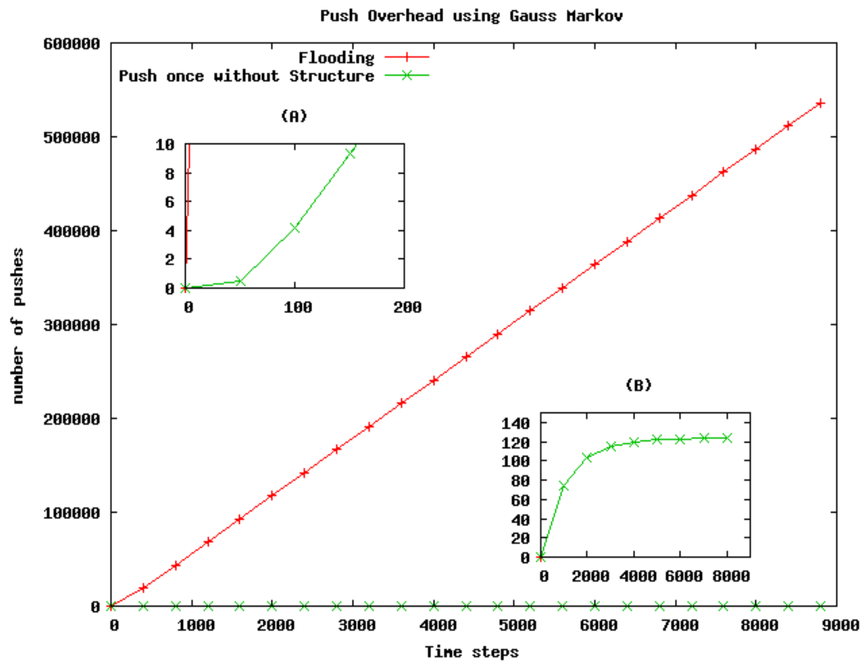


Figure 6.77. Average Push Overhead Costs for Push Once without Structure using Gauss Markov

POWOS only pushes information to nodes that it has never pushed to before. Therefore when the same nodes come in range again, no push of information is occurring. This is different in flooding, where nodes push blindly at every time steps which increase the overhead costs in flooding.

Looking at Figure 6.78, there is a big gap in marginal overhead costs between POWOS

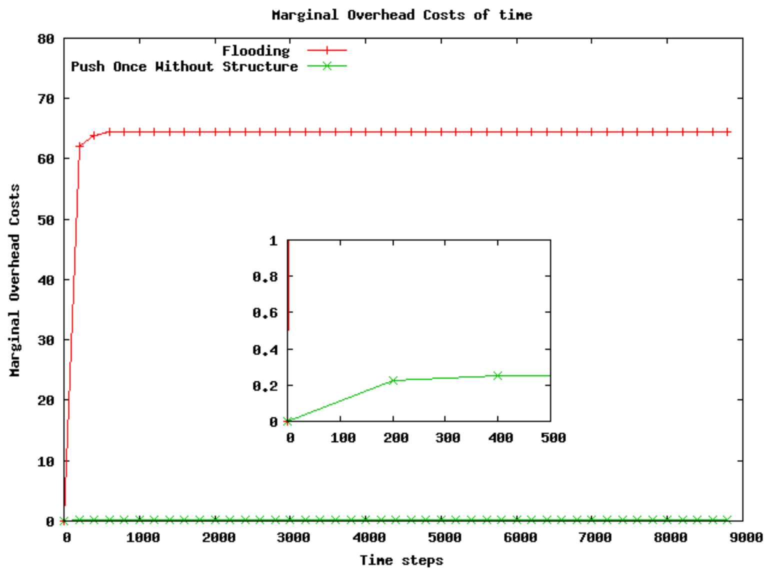


Figure 6.78. Marginal Overhead Costs of time for POWOS using Gauss Markov

and flooding. This is because in POWOS not many nodes are involved in pushing information as nodes are only limited to pushing only to nodes that they never have pushed before. Therefore, the change in marginal overhead cost over time is very small. This is can be seen from the small scale graph in Figure 6.78. However in flooding the marginal overhead costs accelerate at the beginning because of more nodes are become started to actively forward information as the percentage of nodes with an artifact increases over time quickly.

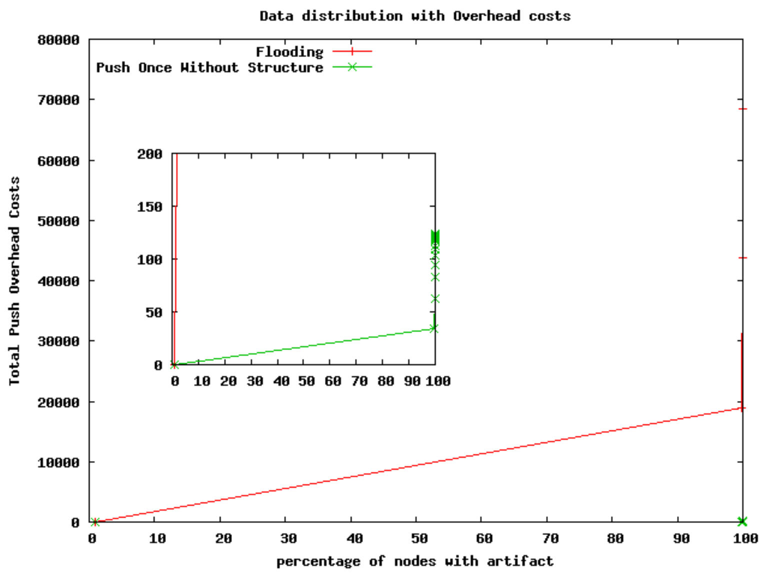


Figure 6.79. Overhead Costs vs Node with artifact for POWOS using GM

Based on Figure 6.79, we can see that there is a big gap in overhead costs and percentage of node with artifact between flooding and POWOS. This is because POWOS uses a very small amount of overhead costs to make the information available quickly as flooding. We can say that POWOS is far more efficient than flooding in managing the cost overhead. Besides the push mechanism, the acceleration of availability of information is also influenced by the use of different mobility models. This is because different mobility model creates different pattern of nodes interactions. The Guass Markov mobility model creates more chance for nodes to interact at the early stage as compared to Random Waypoint. Therefore, the acceleration of cost and percentage of nodes with artifact in Figure 6.79 (Gauss Markov) is different to that in Figure 6.74 (Random Waypoint).

6.6.4.4 Push Once without structure using D-GM

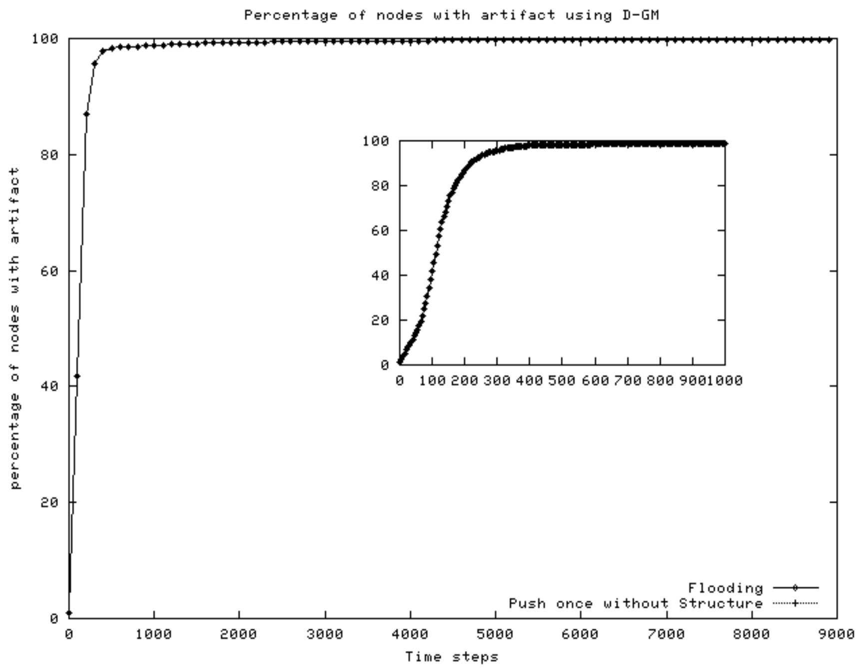


Figure 6.80. Information Profile for Push Once without Structure using D-GM

Figure 6.80 indicates that the POWOS has the same information spreading performance as in the flooding approach. This is because POWOS forwards information the same as flooding but it omits the push duplication in flooding. The duplication is avoided by forwarding information only to the nodes that have never pushed to before. Through this mechanism, POWOS manage reduces the number of pushes involved at ever steps and also able to maintain the percentage of nodes that discover information performance as in the flooding approach.

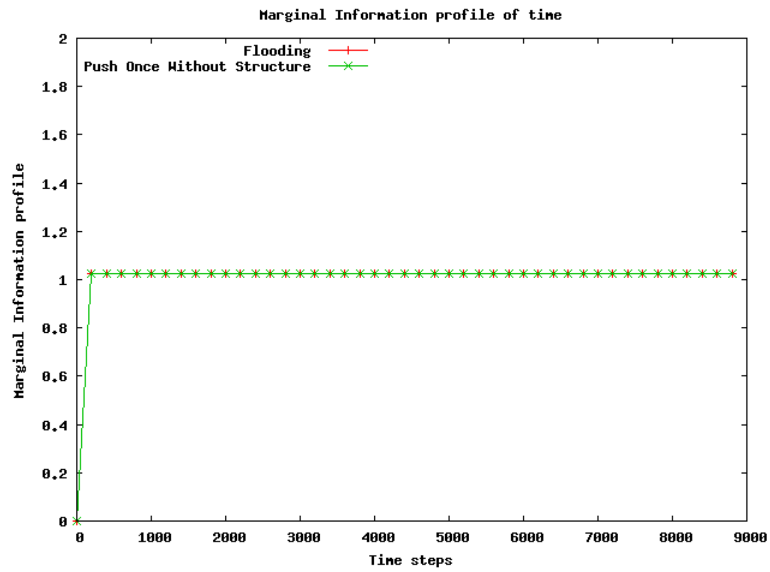


Figure 6.81. Marginal Information profile of time for POWOS using D-GM

Figure 6.81 shows the percentage of nodes with an artifact when a unit of simulation time is increased in both approaches (POWOS and flooding). As we expected POWOS has the same marginal performance as flooding.

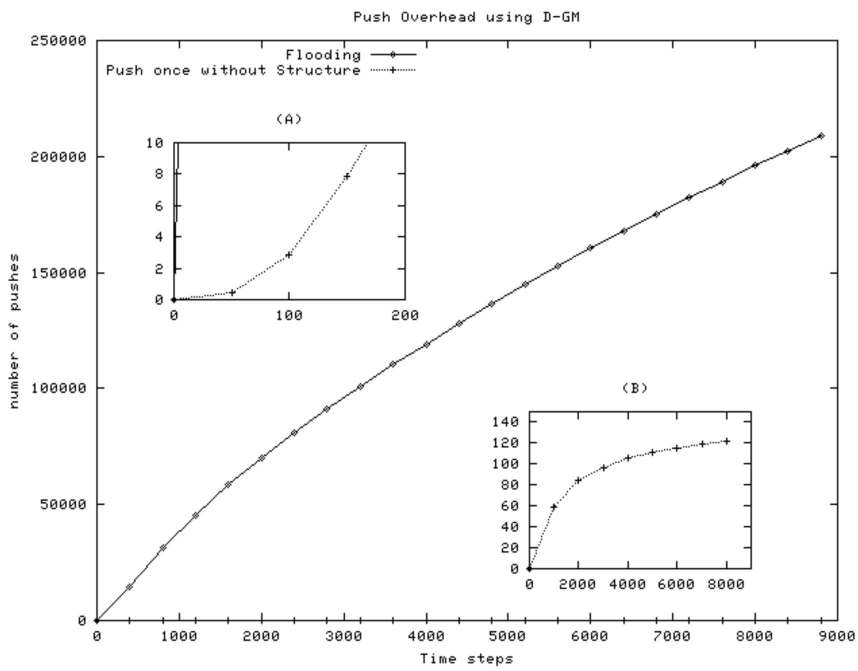


Figure 6.82. Average Push Overhead Costs for Push Once without Structure using D-GM

As in the Gauss Markov mobility model, POWOS with D-GM mobility model has small overhead costs in comparison to flooding. This is can be observed from Figure 6.82 where there is a big gap between flooding and POWOS. In terms of overhead cost, this is because POWOS only pushes information to the nodes that have not pushed to before while in flooding the information is pushed at any meeting opportunity. Therefore more overhead costs incur in flooding approach.

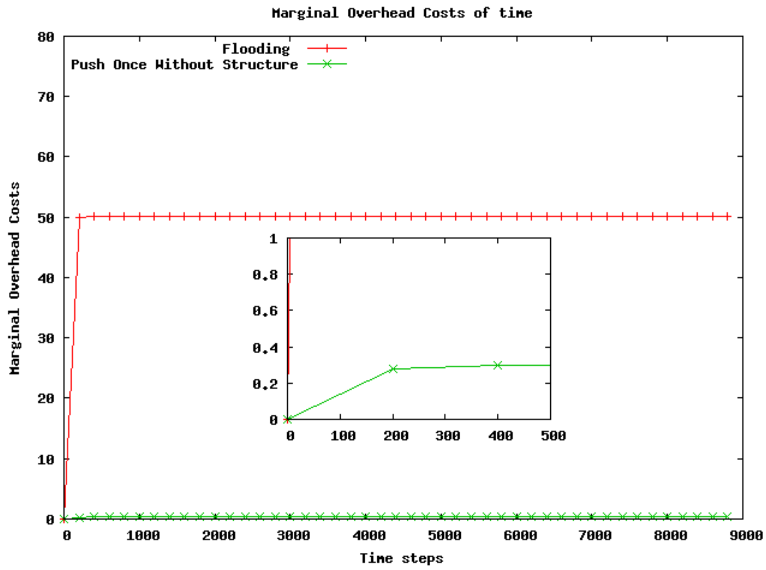


Figure 6.83. Marginal Overhead Costs of time for POWOS using D-GM

Looking at Figure 6.83, the big gap in marginal overhead costs between POWOS and flooding is because in POWOS, not many nodes are involved in pushing information at every time step. This is because nodes that use POWOS only push information to nodes that it has never pushed to before. Therefore, the change of the marginal overhead costs in time is very small. However in the flooding approach the marginal overhead costs accelerates at the beginning because the number of nodes that discover an artifact is increased over time. This situation also increasing the number of nodes that forwarding information over the time.

Based on Figure 6.84, we can see that there is big gap between flooding and POWOS when looking at the change in cost over the performance. Because POWOS and flooding have the same performance, in terms of structure we can say that POWOS is more efficient than flooding in disseminating information in this situation. Beside the efficiency of POWOS, the acceleration of information availability is also influenced by the use of dif-

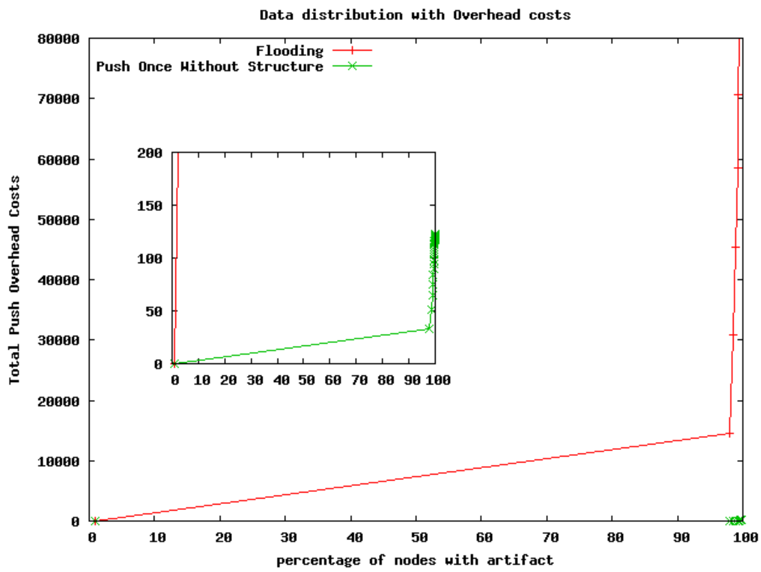


Figure 6.84. Overhead Costs vs Node with artifact for POWOS using D-GM

ferent mobility models. This is because different mobility model creates different pattern of nodes interactions. From our observation, Gauss Markov is better than D-GM mobility model in terms of providing nodes interaction frequency.

6.6.5 Results-Push Once with structure

6.6.5.1 Push Once with structure (POWS) using Random Walk

The differences between POWOS and POWS is the way a peer is selected for forwarding. POWS uses a social structure to guide nodes and to choose its peers. Therefore, in comparing POWS with flooding performance, POWS has a slightly lower performance than flooding. This is because it is required to discover peers that has established a relationship with before more frequently. So, this slightly impedes the information availability performance in POWS. Figure 6.85 shows the performance of POWS in comparison to flooding. The POWS has slightly lower performance compared to flooding because the ability of POWS to push information is subject to the social structure identification.

In Figure 6.86, the POWS marginal costs accelerates close to flooding at the very early unit of time. However after 500 time steps, the marginal gap between POWS and flooding becomes obvious. This is because not many nodes are involved in pushing information as they are constrained by the social structure relationship. So, even though the nodes

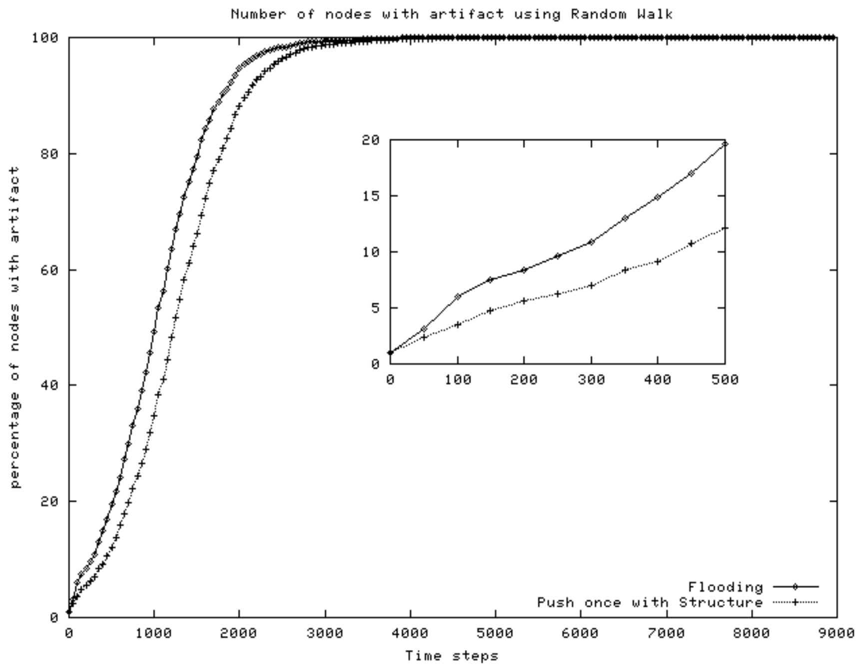


Figure 6.85. Information Profile for Push Once with Structure using Random Walk

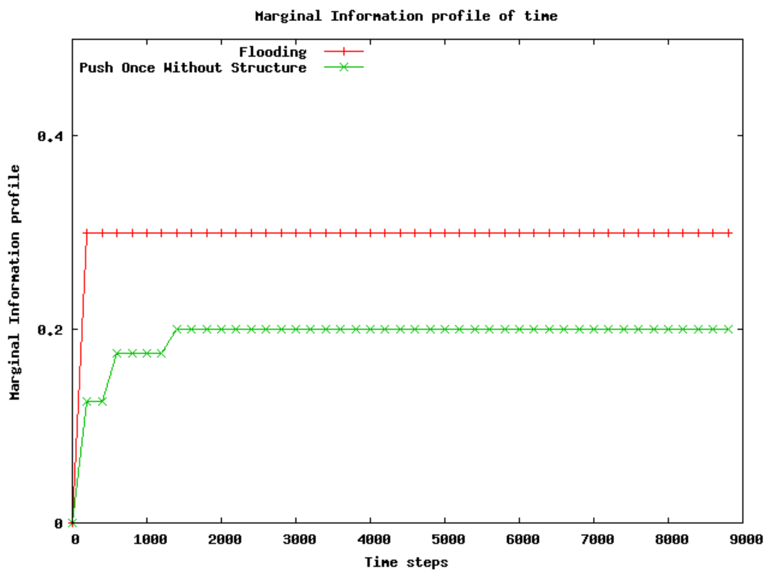


Figure 6.86. Marginal Information profile of time for POWS using Random Walk

that are in range have data to push, a social structure is required to channel effort for pushing of content. This process holds back the performance in POWS as compared to flooding. This is explains why after 500 time steps the marginal information profile of time is decreases when one unit of time is increased.

In Figure 6.87, the total overhead cost in POWS is very small compared to the flooding

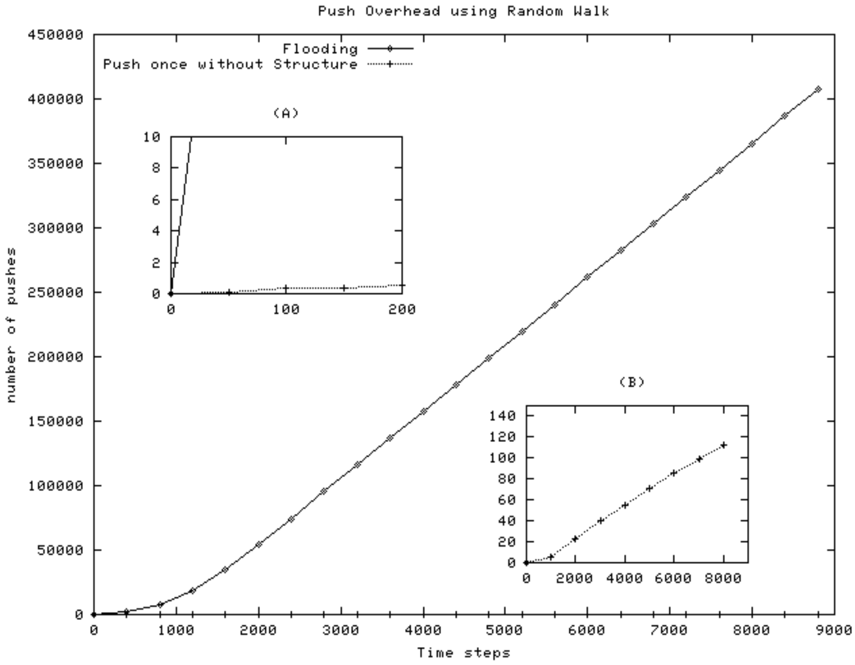


Figure 6.87. Average Push Overhead Costs for POWS using Random Walk

in total overhead costs. This is due to the fact that nodes in POWS only push when co-located with nodes that are listed in the social structure and they have not pushed to before. This mechanism limits push activity in POWS in which minimize the amount of overhead costs in POWS.

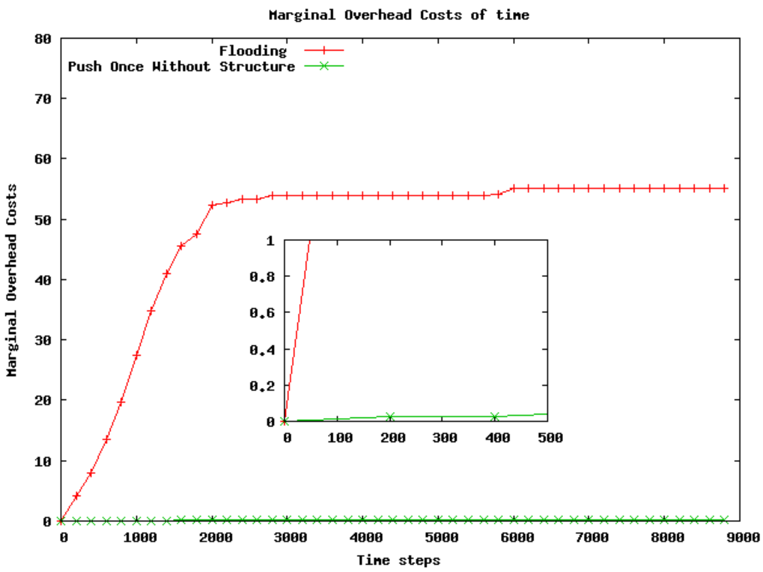


Figure 6.88. Marginal Overhead Costs of time for POWS using Random Walk

From Figure 6.88 we can observe that there is big gap in marginal overhead costs between POWS and flooding. This is because not many nodes in POWS are involved in pushing information as nodes only pushes information to a node that it has never forwarded information to before. In addition, the nodes must be listed in the nodes social structure, otherwise no further forwarding of information is takes place.

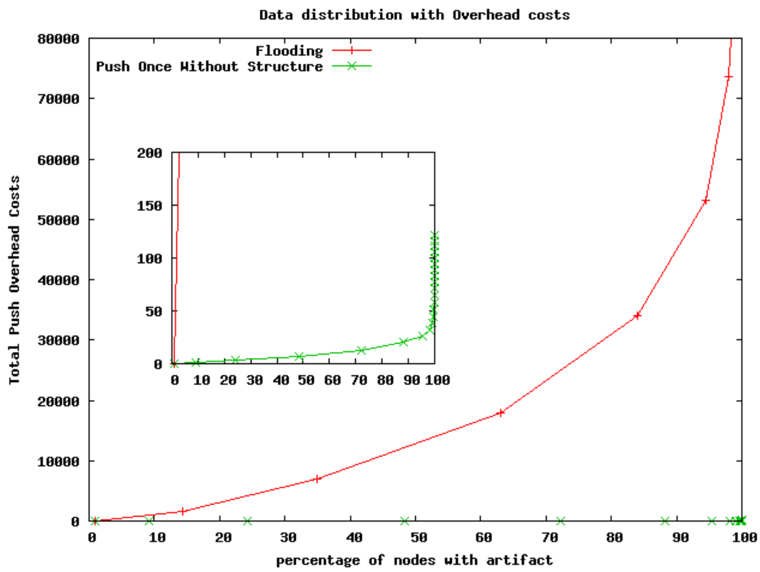


Figure 6.89. Overhead Costs vs Node with artifact for POWS using Random Walk

Looking at Figure 6.89, we can observe that there is a big gap in overhead costs over the percentage of nodes with an artifact between flooding and POWS. This shows that POWS has the ability to control overhead costs that are introduced by flooding. However it still suffers from a delay in making the information available to all nodes as in flooding. Apart from the POWS mechanism, a Random Walk mobility model also effects the information spreading performance and overhead costs that are presented in Figure 6.89. This is because the Random Walk model influences frequency of interactions between nodes.

6.6.5.2 Push Once with structure using Random Waypoint

As we can observe from Figure 6.90, POWS has a slightly lower performance than flooding. This is because the POWS forwarding system is constrained by the social structure relationship. From the POWOS experimentation results, POWOS has a better performance as compared to flooding approach. However, when Push Once is combine with a social structure (POWS), the performance is slightly decreased as more constraints are applied in pushing information to the peers. Therefore, more time is needed to make information

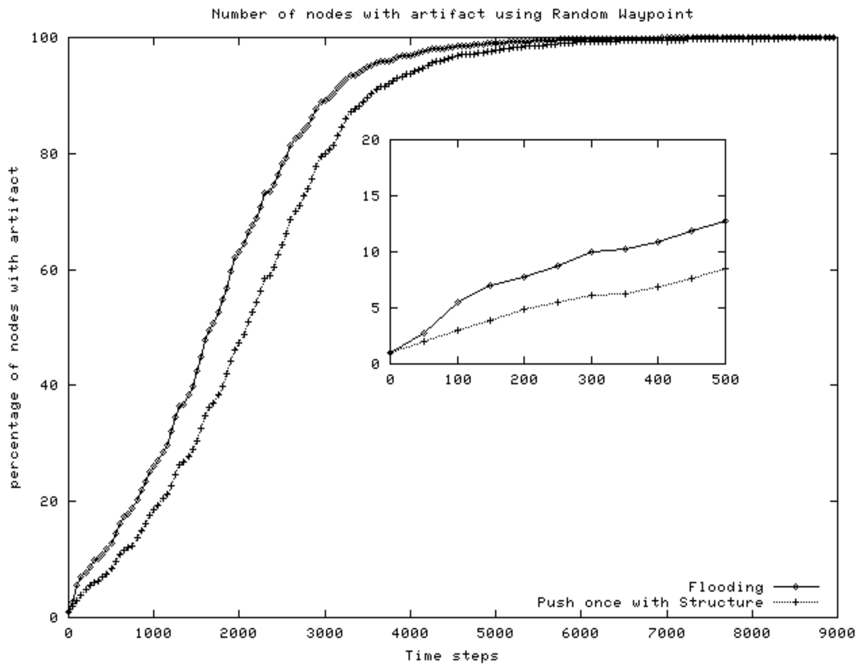


Figure 6.90. Information Profile for Push Once with Structure using Random Waypoint

available to all nodes in the network.

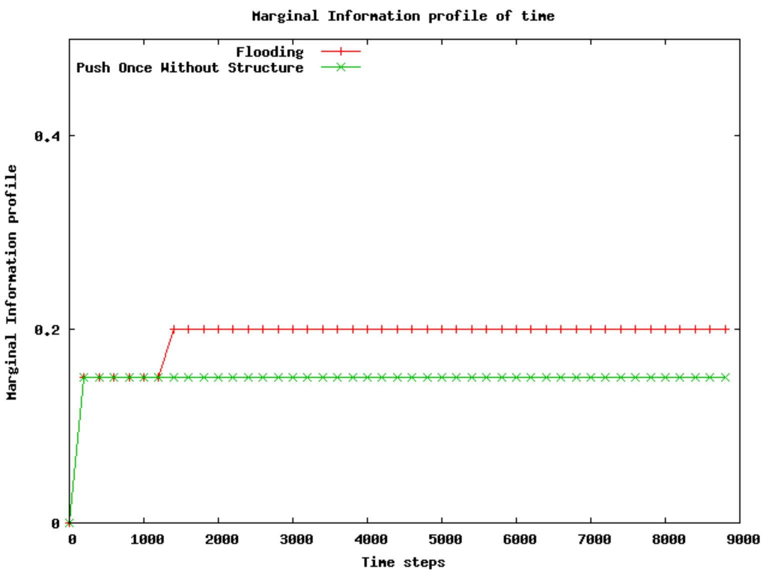


Figure 6.91. Marginal Information profile of time for POWS using Random Waypoint

In Figure 6.91, the POWS marginal information profile over time accelerates at the very early stage. At this stage the nodes in the Random Waypoint model move similar to the Random Walk model, thus the rate of nodes discovering information is similar to the Random Walk model. However after 300 time steps, we observe that the marginal information profile is constant. This shows that the change in number of nodes that are discovering information becomes linearly increasing.

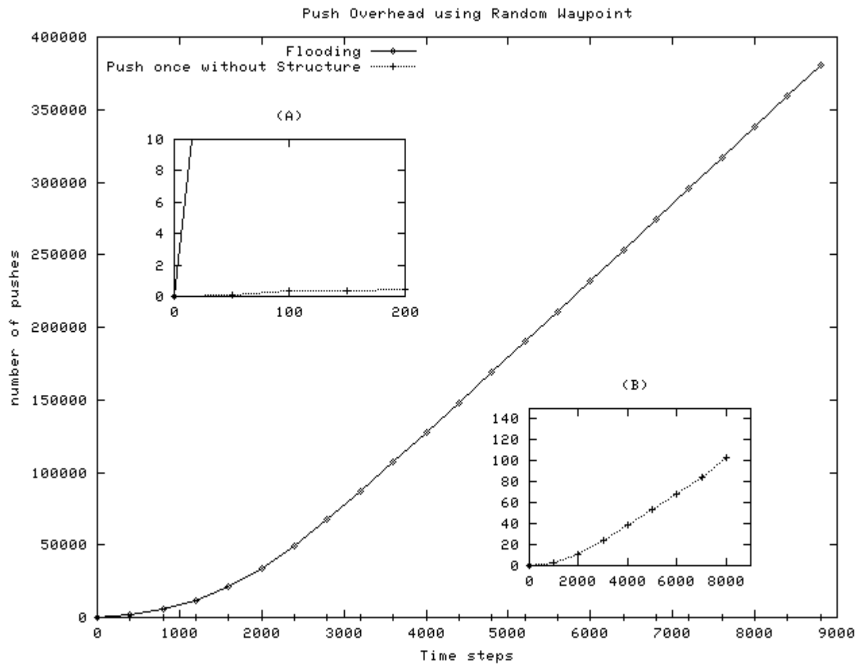


Figure 6.92. Average Push Overhead Costs for Push Once with Structure using Random Waypoint

In terms of overhead cost, POWS has a small amount of overhead costs as not much push activity is involved. In POWS, besides the nodes have to push to the nodes that they never push to before, the nodes also need to make sure that the nodes are listed in social structure relationship. Therefore, there is a very small amount of overhead costs found using POWS.

From Figure 6.93 we can observe that there is a big difference in the marginal overhead costs between POWS and flooding. This is because not many nodes in POWS are involved in pushing information as nodes only push information to the nodes that it has never forwarded information to before. In addition, the nodes also must be in listed in a node's social structure before information can be forwarded. These rules minimize the amount of overhead costs in POWS.

Looking at Figure 6.94, we can observe that there is big gap in overhead costs over percentage of node with artifact between flooding and POWS. POWS has a small overhead costs compared to flooding, however it has an amount of delay in making the information

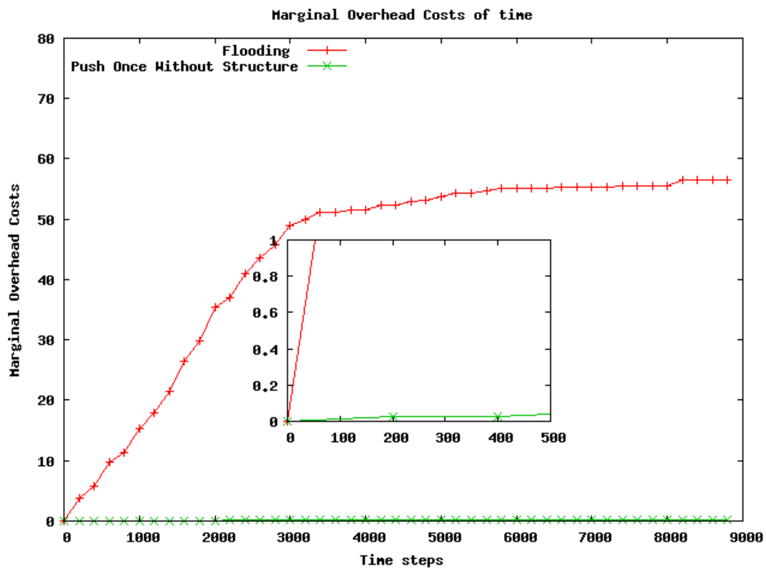


Figure 6.93. Marginal Overhead Costs of time for POWS using Random Waypoint

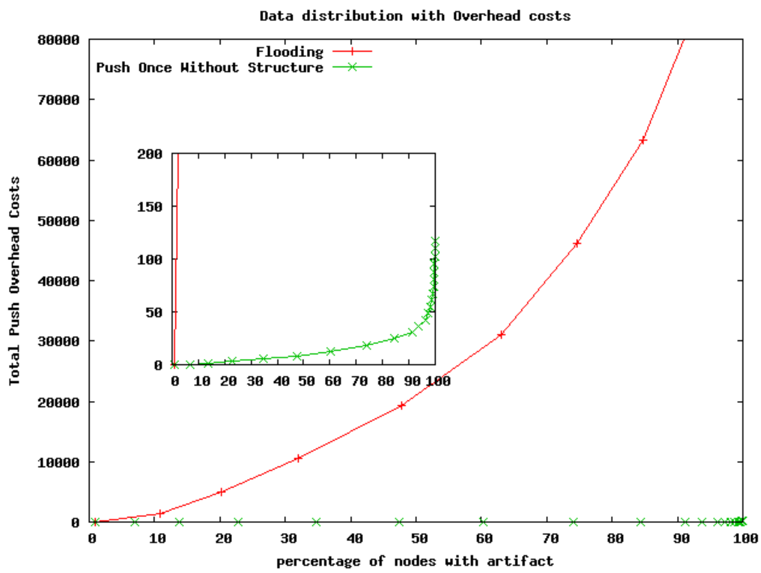


Figure 6.94. Overhead Costs vs Node with artifact for POWS using Random Walk

available to all nodes as compared to flooding. Apart from the POWS mechanism, the Random Waypoint mobility model also influences the information spreading performance and overhead costs as shown in Figure 6.89. This is because Random Waypoint affects the formation of the social structure which determines the frequency of interactions between nodes.

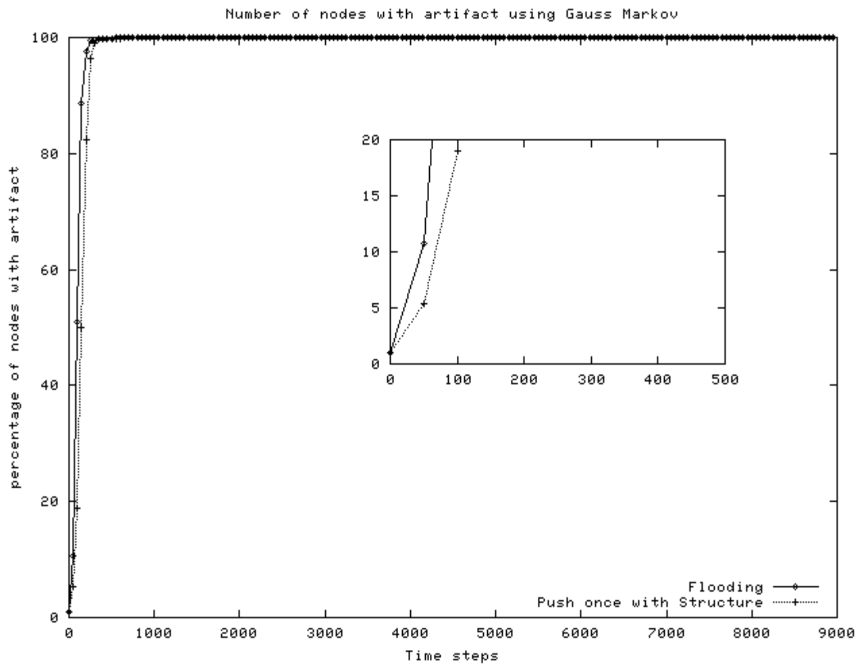


Figure 6.95. Information Profile for Push Once with Structure using Gauss Markov

6.6.5.3 Push Once with structure using Guass Markov

Information spreads very quickly in POWS using the Gauss Markov mobility model as compared to other mobility models. This is due to the fact that the Gauss Markov mobility model creates more chances for nodes to meet different nodes more frequently. From Figure 6.95 we can observe that the POWS is slightly lower that flooding information spreading performance. This is because the push in POWS is dependent on social links. Thus, even though the Gauss Markov offers more frequency of node interactions it has to form the social structure which slows down the POWS information dissemination performance.

Figure 6.96 shows the change in the percentage of nodes with an artifact when one unit of time is increased for both approaches (POWS and flooding). The marginal information profile of time for both approaches accelerate very quickly at the early stage. Looking at the figure, there is a big gap between POWS and flooding. This is due to the fact that the frequency of interaction in POWS are controlled by the social structure apart from the Push Once concept. These factors affect the number of nodes that discover information when one additional unit of time is increased.

In Figure 6.97, the total overhead cost in POWS is very small compared to the flooding

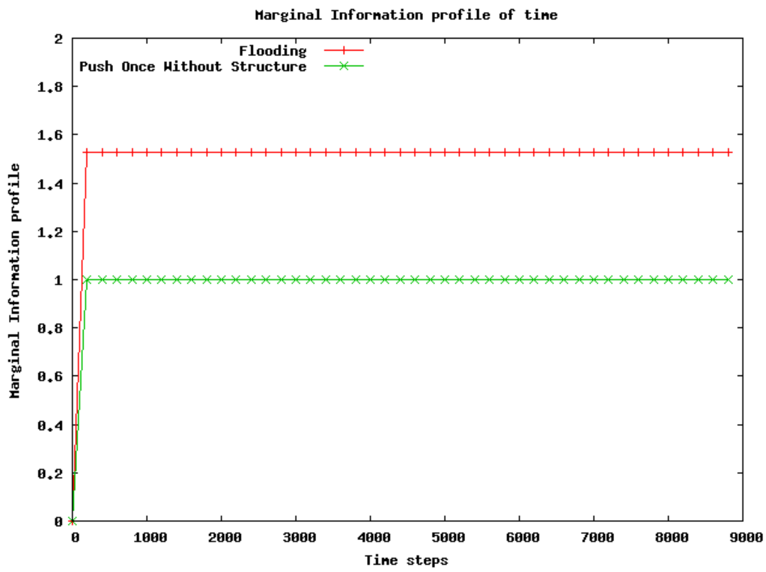


Figure 6.96. Marginal Information profile of time for POWS using GM

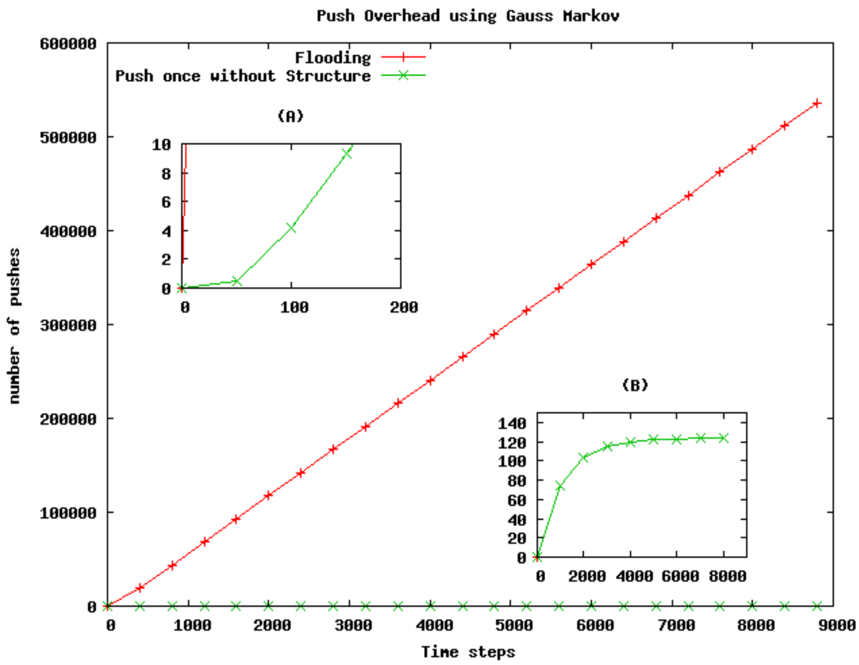


Figure 6.97. Average Push Overhead Costs for Push Once with Structure using Gauss Markov

approach. This is because nodes in POWS only use push resources when nodes are in range with other nodes that is listed in social structure and have not discover an artifact before. Therefore, as we expect there are a small amount of overhead costs in POWS as nodes have more restriction on resource usage.

From Figure 6.98 we can observe that there is big gap in marginal overhead costs

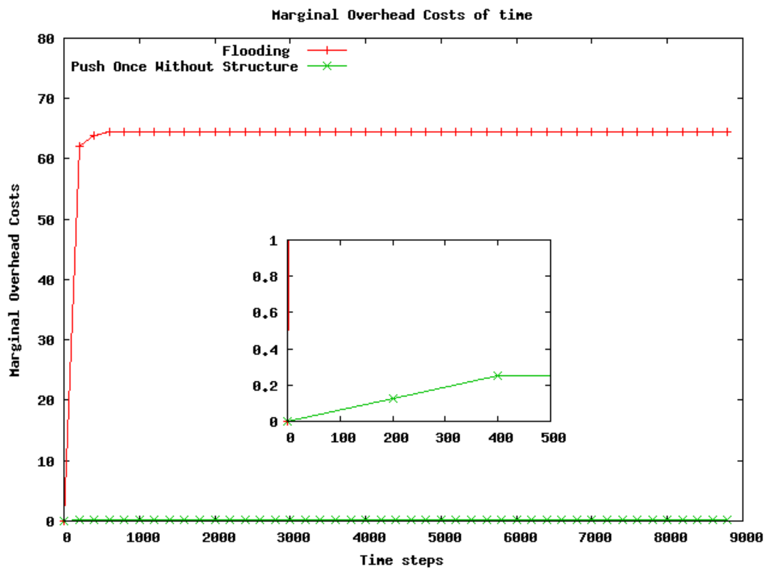


Figure 6.98. Marginal Overhead Costs of time for POWS using Gauss Markov

between POWS and flooding. This is because not many nodes in POWS are involved in pushing information as nodes only push information once. In addition, the nodes also must be in listed in a node's social structure before forwarding information proceeds. These constraints control the push mechanism which results in a small amount of overhead costs being incurred in POWS.

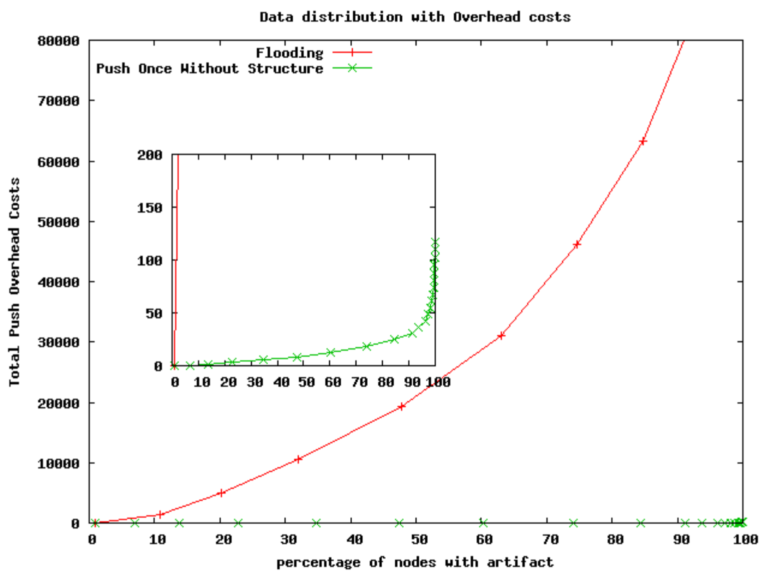


Figure 6.99. Overhead Costs vs Node with artifact for POWS using Random Walk

In Figure 6.99, we can observe that there is big gap in overhead costs with artifacts between flooding and POWS. POWS has a small amount of overhead costs compared to flooding, but it needs more time to disseminate information to all nodes as compared to flooding. Apart from the POWS mechanism, Guass Markov mobility model also plays an important role in shaping the information profile and overhead costs as presented in Figure 6.89.

6.6.5.4 Push Once with structure using D-GM

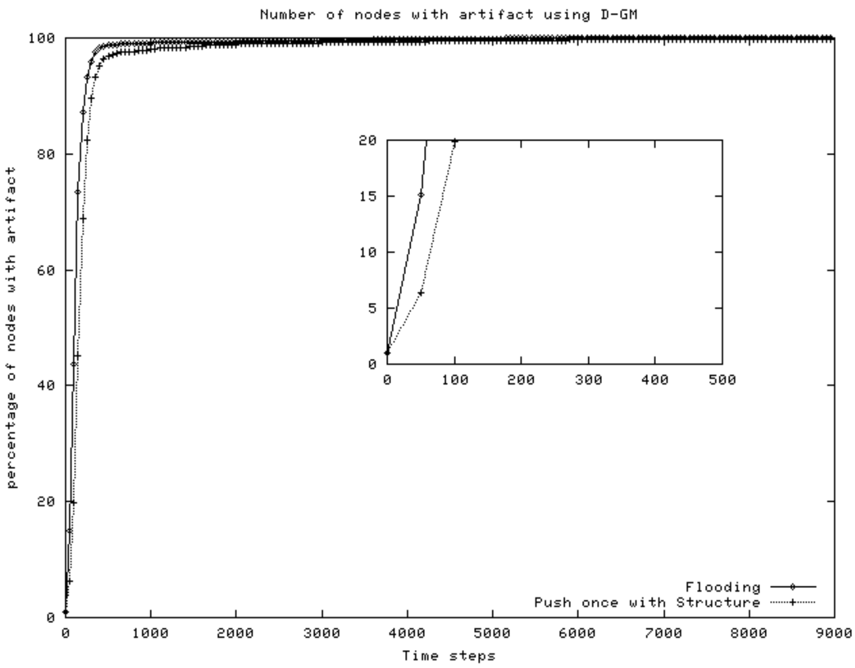


Figure 6.100. Information Profile for Push Once with Structure using D-GM

Using the D-GM mobility model, information spreads quicker than Random Walk but slower than Gauss Markov. This is due to the fact that in the D-GM mobility model nodes have to stop in a particular location before nodes can move to the next location. This mechanism slows down the information spreading process. However, as soon as they are moving, nodes can discover information very quickly. Apart from the mobility model, the information dissemination performance is also influenced by POWS itself. This is because POWS uses a social structure before decided to push information to other nodes. This condition decreases the information dissemination performance as shown in Figure 6.100 relative to flooding.

Based on Figure 6.101 the marginal information profile over time for both approaches accelerates very quickly at the early time stage. This shows that more additional nodes

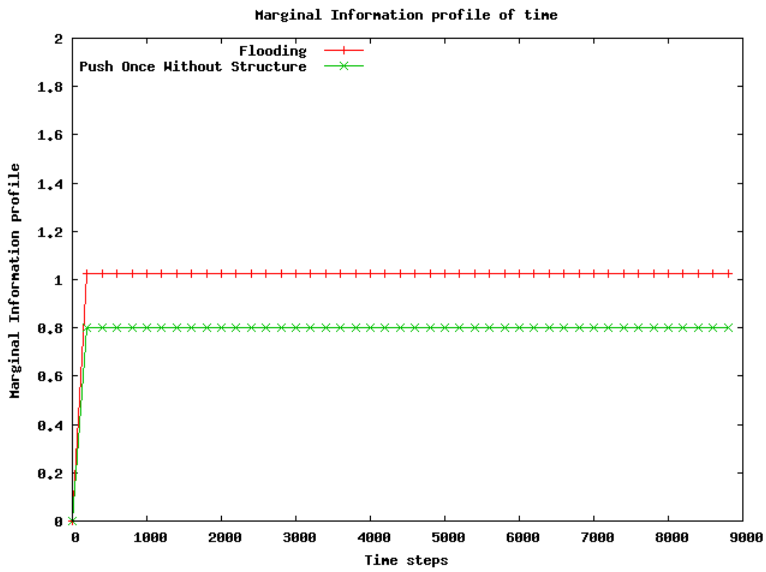


Figure 6.101. Marginal Information profile of time for POWS using D-GM

discover information when a unit of simulation time is increased. From the figure, the gap between POWS and flooding is because of the frequency of interactions in POWS and flooding being different. In POWS the interaction is restricted by the social structure.

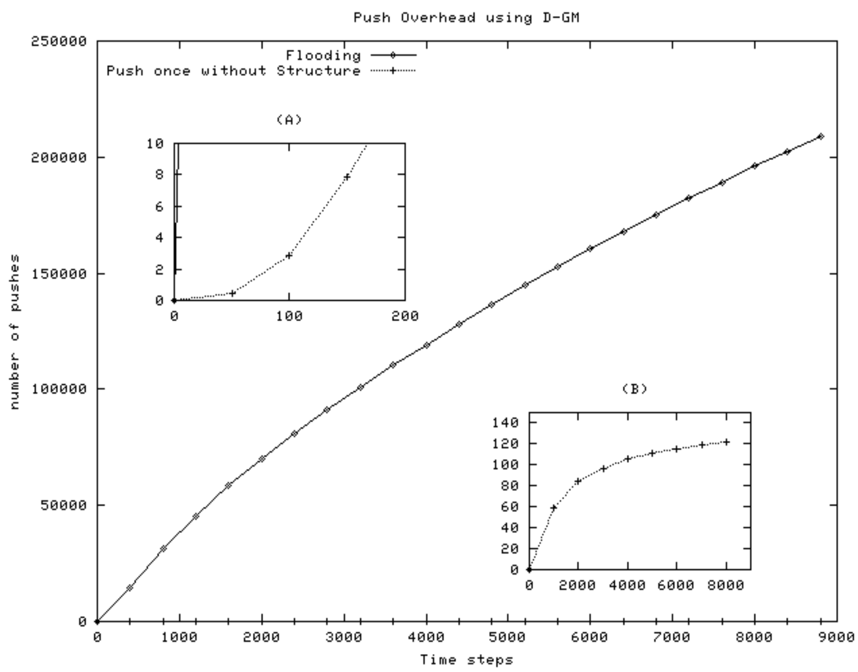


Figure 6.102. Average Push Overhead Costs for Push Once with Structure using D-GM

Looking at the Figure 6.102, the total overhead cost in POWS is very small compared to the flooding total overhead costs. This is due to the fact that nodes in POWS only use the push resource when nodes that are co-located are listed in social structure and have not been forwarded information before. So, this limits the push activity in POWS in which minimize the amount of overhead costs in POWS.

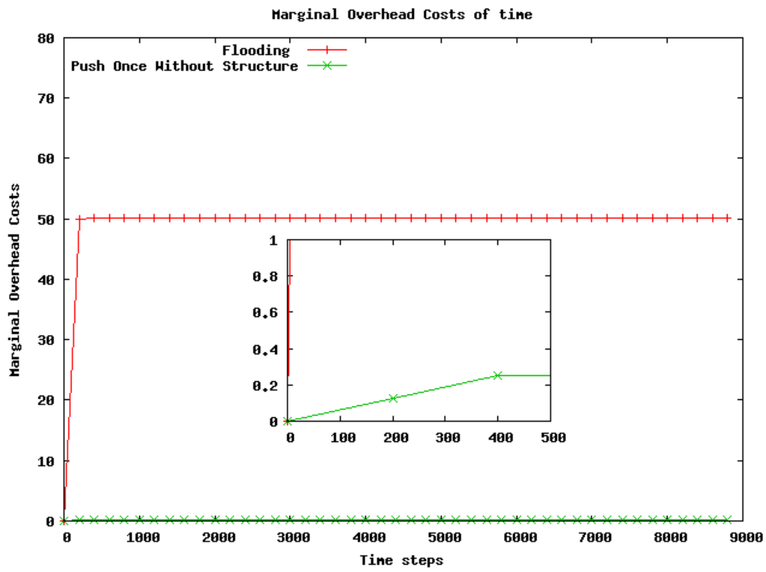


Figure 6.103. Marginal Overhead Costs of time for POWS using D-GM

In Figure 6.103 we can observe that there is big gap in marginal overhead costs between POWS and flooding. In POWS nodes are permitted only to push information to the nodes that it never been push information before. On top of that, the nodes must also listed in nodes SSL.

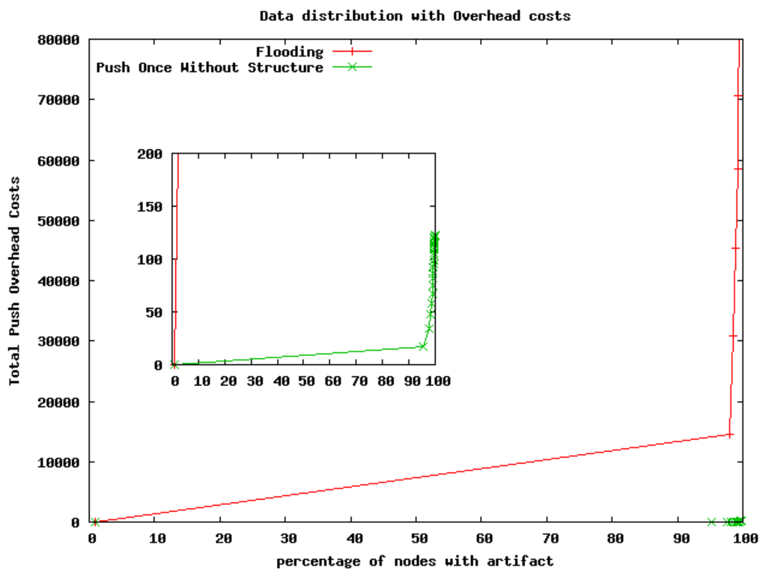


Figure 6.104. Overhead Costs vs Node with artifact for POWS using D-GM

Looking at Figure 6.104, we can observe that there is big gap in overhead costs over percentage of node with artifact between flooding and POWS. This shows that POWS has better management in overhead cost as compared to flooding. However it has delay in making the information available to all nodes. Apart from that, D-GM mobility model also plays an important role in determining the information spreading performance and overhead costs as presented in Figure 6.104.

6.7 Comparison Push Techniques

From the experimentation results, we found that different Push techniques have different impacts on information spreading performance. Pushing is a key process in information spreading. Therefore, modifying the way of pushing information will affect the behavior of information spreading. The statistics in Table 6.6 - 6.10 are taken from Section 6.6.4 and Section 6.6.5 respectively. From the statistics tabulated in the Tables, we observe that, the Gauss Markov mobility model performs better with different push techniques. This is because of Gauss Markov offers more chance for node to meet different nodes at the center of simulation.

Figures 6.105 - 6.108 show different push techniques which are grouped under different mobility models. From the figures, we can see that different push techniques have different information dissemination performance and costs when using different mobility models. This is because mobility models influence a node interaction frequency which indirectly affects the information dissemination coverage and number of push consumed by the node.

From the figures , we can also observe that POWS has the lowest total cost and it has data dissemination performance that close to flooding technique (benchmark performance). This is because POWS manages to overcome the duplication problem in flooding technique through recording the nodes past interactions. Because POWS only pushes information to nodes that it has never met before and are listed in a social structure list, POWS manages to use its push effectively (i.e. it only push information when is really required).

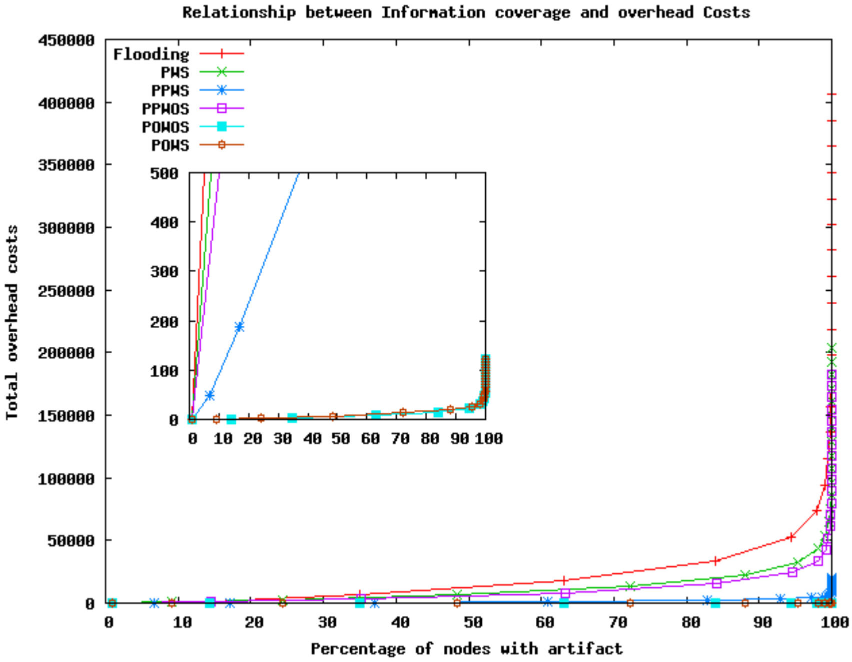


Figure 6.105. Relationship between coverage and costs of each Push technique using **Random Walk** mobility model

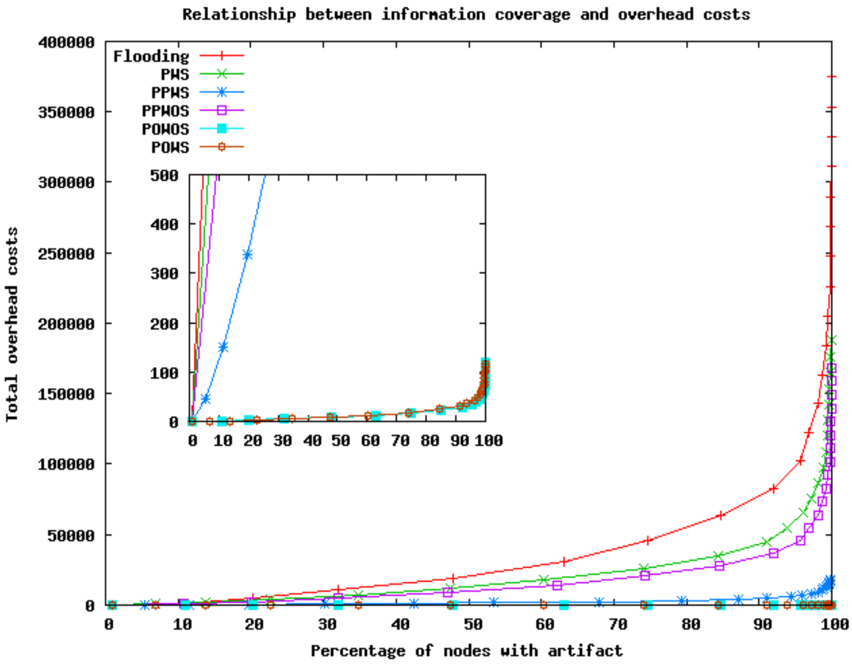


Figure 6.106. Relationship between coverage and costs of each Push technique using **Random Waypoint** mobility model

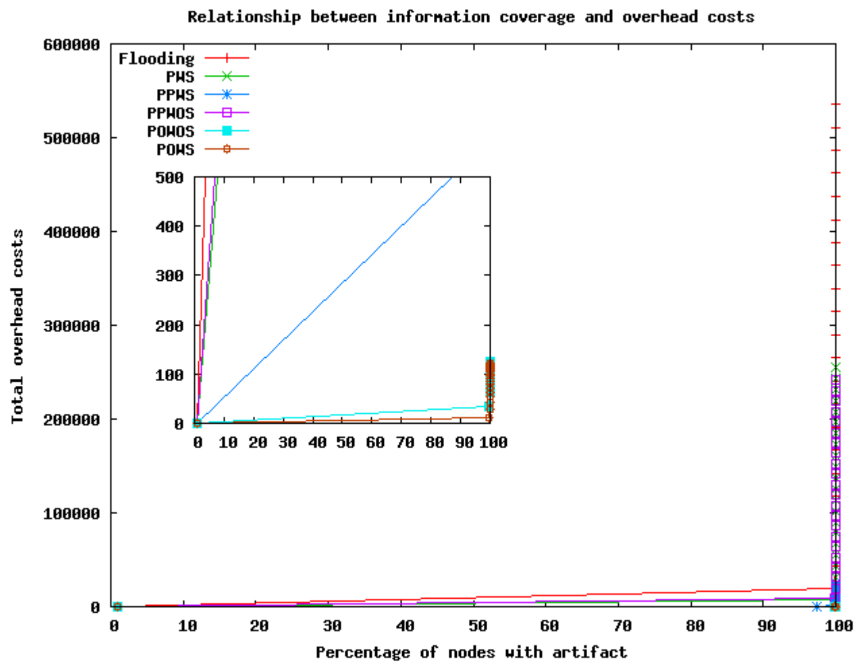


Figure 6.107. Relationship between coverage and costs of each Push technique using **Gauss Markov** mobility model

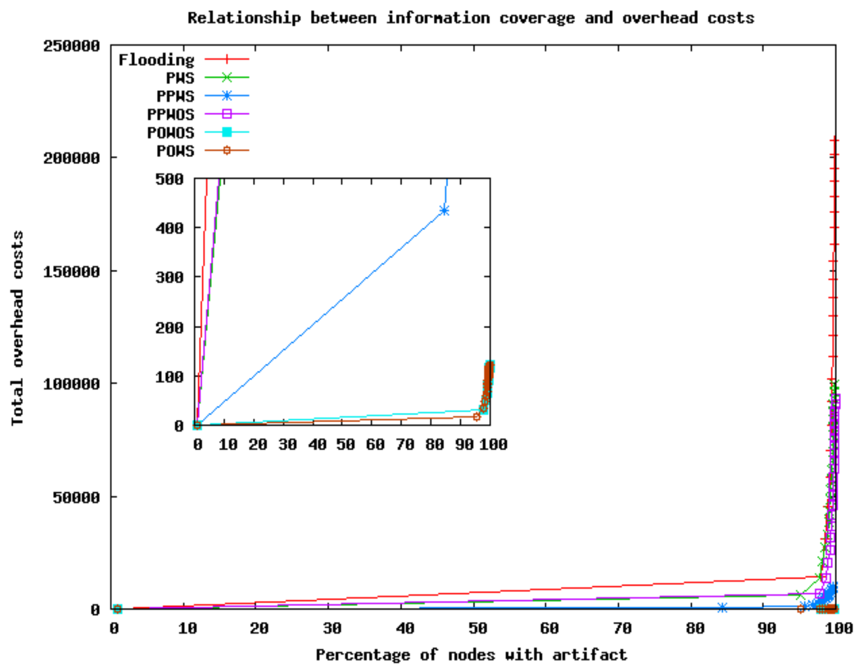


Figure 6.108. Relationship between coverage and costs of each Push technique using **D-GM** mobility model

Table 6.6. Information Profile for different Mobility Models Using **Push Once without structure**

Percentile	*Number of node received an artifact in percentile*			
	Random Walk	Random Waypoint	Gauss Markov	D-GM
5	16.83	12.05	99.84	98.23
15	72.20	37.05	100.00	98.98
25	97.23	71.26	100.00	99.38
50	100.00	98.43	100.00	99.65
75	100.00	100.00	100.00	99.80
100	100.00	100.00	100.00	100.00

Table 6.7. Information Profile for different Mobility Models Using **Push Once with structure**

Percentile	*Number of node received an artifact in percentile*			
	Random Walk	Random Waypoint	Gauss Markov	D-GM
5	11.00	7.83	99.8	96.575
15	58.85	26.85	100.00	98.35
25	93.55	56.65	100.00	99.00
50	100.00	96.70	100.00	99.45
75	100.00	99.48	100.00	99.70
100	100.00	100.00	100.00	100.00

Table 6.8. Information Profile for different Mobility Models Using **Probability Flooding with structure (Probability=0.9)**

Percentile	*Number of node received an artifact in percentile*			
	Random Walk	Random Waypoint	Gauss Markov	D-GM
5	7.65	6.00	99.33	89.28
15	46.63	23.58	100.00	97.25
25	90.58	50.15	100.00	98.10
50	99.95	95.10	100.00	99.00
75	100.00	99.18	100.00	99.40
100	100.00	100.00	100.00	100.00

Table 6.9. Information Profile for different Mobility Models Using **Push Probability without structure (Probability= 0.9)**

Percentile	*Number of node received an artifact in percentile*			
	Random Walk	Random Waypoint	Gauss Markov	D-GM
5	17.03	12.00	99.83	98.175
15	72.55	37.08	100.00	99.03
25	97.33	70.70	100.00	99.36
50	100.00	98.43	100.00	99.73
75	100.00	99.88	100.00	99.88
100	100.00	100.00	100.00	100.00

Table 6.10. Information Profile for different Mobility Models Using **Push with structure**

Percentile	***Number of node received an artifact in percentile***			
	Random Walk	Random Waypoint	Gauss Markov	D-GM
5	11.00	7.83	99.80	96.58
15	58.85	26.85	100.00	98.35
25	93.55	56.65	100.00	99.00
50	100.00	96.7	100.00	99.45
75	100.00	99.48	100.00	99.70
100	100.00	100.00	100.00	100.00

6.8 Conclusion

To recap, the aim of this chapter is to investigate the effect of different push techniques on information spreading performance and its overhead costs. In chapter 4, query and push was investigated and we learnt that it is possible to achieve efficient information dissemination by only focusing on the push mechanism. This is because queries are basically used to guide node to discover information. Furthermore, by omitting query, it also helps to reduce the overhead costs. Chapter 5 showed that social structures can be built up from interactions between nodes. This chapter (chapter 6) has attempted to use these social structures when disseminating information to others, so that overheads are controlled.

From the results presented in this chapter, we found that different ways of pushing information affects the behavior of information dissemination performance. This is because the push mechanism is the key to spreading information. In comparison to Push, Query is only use to discover information. There are three attributes that we have tested with push in this chapter. These are *social structure (structure)*, *probability* and *Push Once (push wisely)*.

The *Push with social structure approach* limits the capability of spreading information quickly. This is because node only pushes information to the node that is listed in its social structure (structure list). So, even though nodes are in range with other nodes (that is not in the social structure), no information is transferred (pushing) between nodes. This is also indicates that a strictness of social structure formulation can decrease the information dissemination as it limits the interactions between nodes that are used for information dissemination. Despite this limitation, it helps in other ways to reduce the number of pushes involved between nodes. This is because, with a social structure, nodes have to push only to nodes that are listed in its social structure.

The *Push with probability approach* is designed to investigate the effect of different levels of push probability on information dissemination performance. We found that increasing the probability level results in high information dissemination performance. This is because when the probability is close to 1, a push occurs at almost every meeting opportunity. So, information easily spreads amongst nodes. However, a high push probability has significant amount of overhead costs.

The *Push only Once approach* is a push approach that uses history of interaction to forward (push) information. From the experimental results, this approach outperforms other push approaches that are introduced in this chapter. This is because it actually performs similar to the flooding technique but avoids pushing information to the same node repeatedly, assisted by the interaction history. Because push information only pushes to nodes that have never been seen before, it has very low overhead costs compared to flooding.

In term of the mobility model, different mobility models have different effects on information dissemination performance. The Gauss Markov model offers better performance compared to other mobility models used in this chapter. This is because Gauss Markov creates more opportunity for nodes to discover each other more frequently. This is indicates that mobility models are important in opportunistic networking and affect information dissemination.

Overall, from this chapter we have investigated the goal of this thesis which is *to minimize the overhead costs* and *at the same time to maintain the information performance as close as possible to flooding technique performance.* Using Social Structure as a mean of disseminating information opportunistic networks is useful in situations where the objective of information spreading is minimizing the use of resources rather than the speed of information availability. However, if the speed of information dissemination is the main concern, then the social structure is not a good approach to be deployed.

Chapter 7

CONCLUSIONS

A key objective of this thesis is to minimize the overhead costs while maintaining the information spreading performance relatively close to flooding performance in opportunistic networks. To achieve the objective we have subdivided the research as follows:

- Investigate the mobility model sensitivity on information dissemination quality through the simulation.
- Design and analyse basic peer to peer interaction protocols.
- Improve the basic peer to peer interaction protocol with intelligent attributes, based around peers building a social structure, which is used to control flooding.

The chapters in this thesis explore these issues. The contributions of each chapter are summarized in section 7.1. The lessons learnt is presented in section 7.2. The potential extension of this research is presented in section 7.3.

7.1 Thesis Contributions

Besides the introduction, literature and conclusion chapters, the following chapters outline the main contributions of this thesis.

- In chapter 3: **the sensitivity of different mobility models with different node density on information dissemination quality are investigated via the simulation**. Four different mobility models (Random Walk, Random Waypoint, Gauss Markov and D-GM) are tested and analyzed. Furthermore the effect of different node density is also analyzed with different mobility models. We discovered that information dissemination in mobile peer to peer interactions is sensitive to the mobility model and the node density used. We also discover that because of limited controls on dissemination, more unnecessary duplications occur. This inspired us to investigate and design the mobile peer to peer interaction protocol in opportunistic networks.

- In chapter 4: **the baseline data dissemination for mobile peer to peer interaction protocol is designed and tested**. Four protocols have been introduced in this chapter, there are called Pure Push, Greedy, L-Push and Spray and Relay. We found that push and query aggressively incur a high consumption of resources. However, the optimal number of pushes per node is difficult to determine because it is subject to the density of the nodes in a particular area. This observation inspired us to investigate whether a social structure method can help to push information efficiently among nodes, by controlled flooding of information.

- In chapter 5: **we proposed three different approaches to investigate whether it is possible to form a social structure through a node interaction with others**. We have designed three approaches to form nodes social structure. These are social structures based on average frequency of interactions, periodicity of interactions and a sliding window. We found that social structure based on a sliding window is the most suitable to be used in a mobile peer to peer scenario, based on the mobility models that we have employed. This is because the construction of the social structure is constructed dynamically according to current nodes interactions. This social structure method is used in Chapter 6 to investigate application of dissemination technique for chapter 4.

- In chapter 6: **we designed different ways of pushing information and analyzed its impact on the overhead cost and the information dissemination performance**. In this chapter we have introduced and tested five different push techniques. There are Push with Structure; Push Probability with Structure; Push Probability without Structure; Push Once without Structure; and Push Once with Structure. We have also combined these with different mobility models. We found that using a social structure approach, a node is able to reduce the cost through pushing information to its social network members only. However this will also decrease the data dissemination performance. The Push Once approach is an intelligent push which has the same performance as flooding. On top of that, it has very small amount of cost in comparison to the flooding approach. With this discovery we are able to make recommendation about the information dissemination performance that is as close as to flooding performance.

As a conclusion, information dissemination in opportunistic networks is possible to be delivered with reduced costs while maintaining the speed of information dissemination performance close to flooding technique. This is can be achieved via an intelligent information dissemination protocol as tested with a number of mobility models. The overhead costs can be further minimised through a social structure approach. However, the speed of information availability to other mobile nodes is slower compared to flooding approach.

7.2 Recommendation

In this section we provide several recommendations based on what we have learned from this research. The recommendations are as follows:

- The density of nodes influences quickness of information availability. For high node density, nodes easily find different potential nodes to relay the information to the final destination. In contrast, with low node density, a better mobility model selection must be in operation to create an opportunity for all nodes to interact with others. This is important because the interactions between nodes is the main condition to spread the information in an opportunistic networks.
- From a mobility model perspective, before investigating data dissemination in opportunistic network, a careful selection of mobility models is required. This is because the information exchange between nodes is dependent on the frequency of nodes seeing each other. High interaction between nodes spreads the information quickly among the nodes. A Random walk mobility model is able to model the random of nodes movement, Random Waypoint is able to model the stop and move of human movement and Gauss Markov mobility model is close to human movement compared

to Random walk and Random Waypoint mobility models movement because it can represent different directions to reach the final destination. D-GM model takes the advantages of Random Waypoint and Gauss Markov mobility models which makes this model more realistic as compared to the other mobility models tested in this research.

- Besides the mobility model, the information exchange protocol also influences the performance of data dissemination. From this research we know that push technique is very effective in disseminating information in which no network infrastructure is exists. However, pushing information blindly has a high consumption of resources and information duplication. So, in order to achieve a good performance with push technique, nodes history of interaction can be used to reduce the overhead of resources.

  The query technique is useful in discovering information. However, without a proper management of querying, a node has high potential to query the same nodes which also contributes to the overuse of resources. To avoid this, the history of the node (past interaction) can be used as one of the tools to utilize the query technique.

  Using the push technique with quota has potential to improve the data dissemination performance. But to get a good performance, the right quota and the size of experiment must be known first. This makes this approach impractical in opportunistic networks when the network topology and density is unknown and uncertain.

- In opportunistic networks, a logical social network exists which can be created from the past nodes interactions history. However, this creation of a social network is dependent on many factors (the nodes mobility models, the nodes social policy interactions, the node frequency interactions, and so forth). The social structure formed with a very strict policy suffers from a high delay of information delivery. This is because the nodes are more selective or exclusive to establish the interactions. However resources are controlled.

  Using social structure as a mean of disseminating information in opportunistic networks is useful in condition where the objective of information spreading is focus on minimizing the use of resources rather than the speed of information availability.

However, if the speed of information dissemination is the priority objective, then the social structure in not a good option to be used.

- **Push information intelligently** is a very efficient technique in spreading information via opportunistic networks. By keeping track on whom the node has pushed information to, not only can duplication be avoided but it also ensures that each node that is in range receives information without giving any information about their identity. However this technique requires a sufficient memory storage to remember the past node interactions.

In summary, from this research, in order to be able to achieve a good data dissemination performance in opportunistic networks, the following issues need to be considered:

- The node density.
- The placement of information Source (if required).
- The use of mobility models.
- The information exchange protocols.
- The use of the nodes past interactions history.
- The importance of speed of dissemination verses resource usage.

Besides the performance issues mentioned above, there are more overheads (i.e. node discovery, memory management, lookup table) that need to be considered to measure the data dissemination performance in practical opportunistic networks which are not addressed in this thesis.

7.3 Future Works

This research paves the way for several potential research areas in improving the data dissemination in opportunistic networks. The following list potential research that can be conducted as an extension of works from this research.

- Since this research focuses on homogeneous information, the next potential research is looking at heterogeneous information. Investigating heterogenous information involves profiling the nodes preferences. Therefore a lot of overhead cost in maintaining the accuracy of profile as to helps nodes to improve the quality of information exchange.

- Reliability and accuracy of information is critical in information dissemination. Basically the reliability and accuracy of information is dependent on time and location. However, investigating these issues require a better synchronization with the user profile. This is because user profiles determine information that are related to the particular consumer (receiver).

- Using an opportunistic networking platform to disseminate information in a small area and measure the effectiveness of different dissemination approaches which are introduced in this thesis is also a vital research issue. However to establish the research, it requires collaboration from users (mobile nodes) to participate as information relays or gateways to allow information dissemination. The advantages of using real application is tested with other hidden factors that affect the data dissemination performance. So, the result can be used further to validate the finding that has be presented in this thesis.

Realizing the work to understand the data dissemination in opportunistic network via simulation requires further work to make the interactions between nodes more reliable and accurate at the operational level.

BIBLIOGRAPHY

[1] K. Aberer, P. Curdre-Maurox, A. Datta, Z. Despotovic, M. Hauswirth, M. punceva, R. Schmidt, and J. Wu. Advanced peer-to-peer networking: The p-grid system and its application. *PIK journal - praxis de informationsverarbeitung und Kommunikation, Special issue on P2P system*, 2003.

[2] D. Agrawal, A. El Abbadi, and R. C. Steinke. Epidemic algorithms in replicated databases (extended abstract). In *PODS '97: Proceedings of the sixteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*, pages 161–172, New York, NY, USA, 1997. ACM.

[3] Robert A.Hanneman and Mark Riddle. Centrality and power. `http://www.faculty.ucr.edu/~hannemen/nettext/C10_Centrality.html`.

[4] Allan Beaufour, Martin Leopold, and Philippe Bonnet. Smart-tag based data dissemination. In *WSNA '02: Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications*, pages 68–77, New York, NY, USA, 2002. ACM.

[5] C. Boldrini, M. Conti, I. Iacopini, and A. Passerella. Hibop: a history based routing protocol for opportunistic networks. In *IEEE Internation symposium World of Wireless and Multimedia Networks*, June 2007.

[6] C. Boldrini, M. Conti, and A. Passarella. Contentplace: Social-aware data dissemination in opportunistic networks. In *MSWiM'08*, Vancouver,BC,Cananda, October 27-31 2008.

[7] Tracy Camp, Jeff Boleng, and Vanessa Davies. A survey of mobility models for ad hoc network research. *Wireless Communications and Mobile Computing*, 2(5):483 – 502, 2002.

[8] A. Datta. Mobigrid: Peer to peer overlay and mobile network rendezvous- a data management perspective. In *CAiSE 2003 Doctoral Sysmposium*, 2003.

[9] Avri Doria. Providing connectivity to the saami nomadic community. In *In Proc. 2nd Int. Conf. on Open Collaborative Design for Sustainable Innovation*, 2002.

[10] Natalie Glance, Dave Snowdon, and Jean-Luc Meunier. Pollen: using people as a communication medium. *Comput. Netw.*, 35(4):429–442, 2001.

[11] Matthias Grossglauser and Martin Vetterli. Locating mobile nodes with ease: learning efficient routes from encounter histories alone. *IEEE/ACM Trans. Netw.*, 14(3):457–469, 2006.

[12] John Guare. Six degrees of separation. `http://en.wikipedia.org/wiki/Six_degrees_of_separation`.

[13] T. Hakkarainen, V. Saviko, and A. Lattunen. Generic engine for collaborative mobile applications. In *IADIS International Conference WWWW/Internet 2005*, pages 243–246, 2005.

[14] Wenbo He, Ying Huang, K. Nahrstedt, and Bo Wu. Message propagation in ad-hoc-based proximity mobile social networks. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2010 8th IEEE International Conference on*, pages 141–146, April 2010.

[15] T. Horozov, A. Grama, V. Vasudevan, and S. Landis. Moby-a mobile peer-to-peer service and data network. In *Parallel Processing, 2002. Proceedings. International Conference on*, pages 437 – 444, 2002.

[16] Sushant Jain, Kevin Fall, and Rabin Patra. Routing in a delay tolerant network. In *SIGCOMM '04: Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 145–158, New York, NY, USA, 2004. ACM.

[17] D.B. Johnson and D.A. Maltz. large version dynamic source routing in ad hoc wireless networks. *Mobile Computing*, 353, 1996.

[18] Evan P. C. Jones, Lily Li, and Paul A. S. Ward. Practical routing in delay-tolerant networks. In *WDTN '05: Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pages 237–243, New York, NY, USA, 2005. ACM.

[19] Philo Juang, Hidekazu Oki, Yong Wang, Margaret Martonosi, Li Shiuan Peh, and Daniel Rubenstein. Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with zebranet. *SIGOPS Oper. Syst. Rev.*, 36(5):96–107, 2002.

[20] JXTA. `http://www.jxta.org`.

[21] D. Knoke and S. Yang. *Social Network Analysis (Quantative Application in the Social Science).* Sage Inc., 2nd edition edition, 2007.

[22] G. Kortuem, J. Schneider, D. Preuitt, T.G.C. Thompson, S. Fickas, and Z. Segall. When peer-to-peer comes face-to-face: collaborative peer-to-peer computing in mobile ad-hoc networks. In *Peer-to-Peer Computing, 2001. Proceedings. First International Conference on*, pages 75 –91, aug 2001.

[23] N. Kotilainen, M. Weber, M. Vapa, and J. Vuori. Mobile chedar - a peer to peer middleware for mobile devices. In *Pervasive Computing abd Communication Workshops, 2005*, pages 86–90, 2005.

[24] Chul-Ho Lee and Do Young Eun. Exploiting heterogeneity in mobile opportunistic networks: An analytic approach. In *Sensor Mesh and Ad Hoc Communications and Networks (SECON), 2010 7th Annual IEEE Communications Society Conference on*, pages 1 –9, June 2010.

[25] H. Lenando and R.M Whitaker. Maintaining spatial-temporal knowledge through human interaction. In *Bio-Inspired Computing and Communication: First Workshop on Bio-Inspired Design of Networks*, pages 307–318, 2008.

[26] V. Lenders, G. Karlsson, and M. May. Wireless ad hoc podcasting. In *IEEE SECON*, 2007.

[27] A. Lindgren, A. Doria, and O. Schelen. Probabilistic routing in intermittently connected networks. In *Lecture Notes in Computer Science*, pages 239–254. Springer Berlin / Heidelberg, 2005.

[28] G. Di Modica M. Bisignano and O. Tomarchio. Jmobipeer a middleware for mobile peer-to-peer computing in manet. In *IEEE Internation Conference on Distributed Computing Systems Workshops*, pages 785–791. IEEE, 2005.

[29] S. Giordano M. Conti, M. May, and A. Passarella. From opportunistic networks to opportunistic computing. *Communications Magazine, IEEE*, 48(9):126 –139, 2010.

[30] M. Matuszewski, N. Beijir, J. Lehtinen, and T. Hyyrylainen. Mobile peer to peer content sharing application. *Consumer Communications and Networking Conference 2006, CCNC IEEE*, 2(8):1324–1325, 2006.

[31] Stanley Milgram. Small world experiment. `http://en.wikipedia.org/wiki/Small_world_experiment`.

[32] M. Musolesi, S. Hailes, and C. Mascolo. Adaptive routing for intermittently connected ad hoc networks. In *WoWMoM '05: Proceeding 6th International Symposium on a World of Wireless, Mobile, and Multimedia Networks*. IEEE Press, 2005.

[33] S.Y Ni, Y.C. Tseng, Y.S Chen, and J.P Sheu. The broadcast storm problem in a mobile ad hoc network. In *Conference on Mobile Computing and Networking (MobilCom'99)*, pages 151–162, 1999.

[34] L. Pelusi, A. Passarella, and Marco Conti. Opportunistic networking: data forwarding in disconnected mobile ad hoc networks. In *Communication Magazine,IEEE*, volume 44, pages 134–141, 2006.

[35] A. Pentland and R. Fletcher. Daknet: Rethingking connectivity in developing nation. In *In IEEE Computer Society*, 2004.

[36] C. Perkins, E. Belding-Royer, and S. Das. Ad hoc on-demand distance vector routing, 2003.

[37] Xiong Yong ping, Limin Sun, Wenbo He, and Jian Ma. Anycast routing in mobile opportunistic networks. In *Computers and Communications (ISCC), 2010 IEEE Symposium on*, pages 599 –604, 2010.

[38] Y. Sasson, D. Cavin, and A. Schiper. Probabilistic broadcast for flooding in wireless mobile ad hoc networks. In *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, volume 2, pages 1124 –1130 vol.2, march 2003.

[39] K. Scoot and S. Burleigh. `http://www.dtnrg.org/docs/specs/draft-irtf-dtnrg-bundle-spec-03`.

[40] R.C. Shah, S. Roy, S. Jain, and W. Brunette. Data mules: Modeling a three-tier for spare sensor network. In *Elsevier Ad Hoc Networks Journal*, volume 1 of *2-3*, pages 215–233, september 2003.

[41] R. Shcollmeire, I. Gruber, and F. Niethmanmer. Protocol for peer-to-peer networking in mobile environment. In *The 12th Internatinal Conference on Commputer Communication and Networks*, pages 121–127. IEEE, 2003.

[42] Tara Small and Zygmunt J. Haas. The shared wireless infostation model: a new ad hoc networking paradigm (or where there is a whale, there is a way). In *MobiHoc '03: Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing*, pages 233–244, New York, NY, USA, 2003. ACM.

[43] T. Spyropoulos, K. Psounis, and C.S. Raghavendra. Spray and wait: An effcient routing scheme for intermittently connected mobile networks. In *In ACM SIGGCOMM Workshop*, Philadelphia.

[44] T. Spyropoulos, K. Psounis, and C.S. Raghavendra. Multiple-copy routing in intermittently connected mobile networks. In *Technical Report CENG-2004-12*, 2004.

[45] T. Spyropoulos, K. Psounis, and C.S. Raghavendra. Single-copy routing in intermittently connected mobile networks. In *Proceeding of IEEEE Secon'04*, 2004.

[46] M.M Bin Tariq, M. Ammar, and E. Zegura. Message ferry route design for sparse ad hoc networks with mobile nodes. In *MobiHoc '06: Proceedings of the 7th ACM international symposium on Mobile ad hoc networking and computing*, pages 37–48, New York, NY, USA, 2006. ACM.

[47] TechTarget. `http://searchnetworking.techtarget.com/definition/unicast`.

[48] A. Vahdat and D. Becker. Epidemic routing for partially connected ad hoc networks. In *Technical Report*, CS Department,Duke University,PA,USA, 2000.

[49] A.I. Wang, T. Bjornsgard, and K. Saxlund. Peer2me - rapid application framework fo mobile peer-to-peer applications. In *the 2007 International symposium on Collaborative Technologies and System*, May.

[50] Wikipedia. Gossip protocol. `http://en.wikipedia.org/wiki/Gossip_protocol`.

[51] R. Winter, T. Zahn, and J. Schiller. Dynamo: A topology-aware p2p overlay network for dynamic, mobile ad-hoc environment. *Telecommunication System*, 27(2):321, 2004.

[52] E. Yoneki, P. Hui, S. Chan, and J. Crowcroft. A socio-aware overlay for publish/subscribe communication in delay tolerent networks. In *ACM MSWiM*, 2007.

[53] W. Zhao, M. Ammar, and E. Zegura. Proactive routing in highly-partitioned wireless and ad hoc networks. In *9th IEEE Workshop on Furture Trends in Distributed Computing System (FTDCS)*, May 2003.

[54] W. Zhao, M. Ammar, and E. Zegura. A message ferrying approach for data delivery in spare mobile ad hoc networks. In *5th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'2004)*, May 2004.

[55] W. Zhao, M. Ammar, and E. Zegura. Controlling the mobility of multiple data transpot ferries in a delay-tolerant network. In *IEEE Infocom*, March 2005.